Mehrkanalige Geräuschreduktion bei Sprachsignalen mittels Kalman-Filter

Dem Fachbereich Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt

zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
vorgelegte

Dissertation

von

Dipl.-Ing. Alexander Michael Kaps

geboren am 03. Juni 1975 in Braunfels

Lindau (Bodensee)
2008

Referent: Prof. Dr.-Ing. E. Hänsler
Korreferent: Prof. Dr.-Ing. P. Meißner

Tag der Einreichung: 11.02.2008
Tag der Prüfung: 30.06.2008

D 17
Darmstädter Dissertation

Vorwort

Die vorliegende Arbeit entstand im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Fachgebiet Theorie der Signale des Instituts für Nachrichtentechnik der Technischen Universität Darmstadt.

Mein besonderer Dank gilt Herrn Prof. Dr.-Ing. E. Hänsler für die Betreuung dieser Arbeit, die zahlreichen fachlichen Diskussionen und die vielfältige Unterstützung, die ich durch ihn erfahren habe. Ebenso danke ich Herrn Prof. Dr.-Ing. P. Meißner für die Übernahme des Korreferats und sein Interesse an meiner Arbeit. Weiterhin möchte ich Frau Prof. Dr.-Ing. A. Klein sowie den Herren Prof. Dr.-Ing. J. Stenzel und Prof. Dr.-Ing. J. Adamy für Ihre Mitwirkung in der Prüfungskommission danken.

Bei der Firma Harman/Becker in Ulm bedanke ich mich für das zur Verfügung Stellen der in dieser Arbeit verwendeten Audiodaten. Dabei gilt mein besonderer Dank Herrn Dr.-Ing. G. Schmidt für seine stets wertvollen Hinweise und Anregungen.

Die Zeit am Fachgebiet war von einem offenen und freundschaftlichen Verhältnis unter den Mitarbeiterinnen und Mitarbeitern geprägt. Für diese sehr angenehme und motivierende Atmosphäre möchte ich allen ehemaligen Kolleginnen und Kollegen danken. Ebenso sei allen Damen und Herren, die im Rahmen von Studien- und Diplomarbeiten Beiträge zu dieser Arbeit geleistet haben, hier gedankt.

Für das Korrekturlesen der Arbeit danke ich den Herren Dr.-Ing. M. Darms, Dr.-Ing. M. Eisenacher und Dipl.-Wirtsch.-Ing. P. Wieber.

Schließlich möchte ich mich bei meiner Familie sowie bei allen Freunden und Bekannten bedanken, die mich auf meinem Weg begleitet und unterstützt haben.

Lindau (Bodensee), im Juli 2008 Alexander Kaps

Inhaltsverzeichnis

Kapitel 1

Einleitung

1.1 Motivation

Freisprecheinrichtungen erhöhen nicht nur den Komfort sondern auch die Sicherheit beim Fahren. Deshalb ist in Deutschland seit Anfang des Jahres 2001 die Verwendung einer Freisprecheinrichtung für das Telefonieren während der Fahrt gesetzlich vorgeschrieben [47]. Dies hat zusammen mit immer leistungsfähigeren und gleichzeitig kostengünstigeren Mikroprozessoren dazu geführt, dass heute eine Vielzahl solcher Systeme zur Verfügung stehen, auf denen immer komplexere Algorithmen implementierbar sind. Dabei werden diese Systeme entweder als fest im Kraftfahrzeug installierte oder portable Geräte angeboten.

Waren fest installierte Systeme zunächst ausschließlich Fahrzeugen der Oberklasse vorbehalten, bieten heute alle deutschen Automobilhersteller solche Systeme ab der Mittel- bzw. Kompaktklasse an. Auch die Ausstattung der Freisprecheinrichtungen hat sich merklich verbessert. „Scheuten [...] noch die meisten Hersteller" Anfang des Jahres 2003, „den Aufwand und die Kosten mehrerer Mikrofone" [39], so sind heute Arrays[1] mit bis zu vier Mikrofonen Standard in Oberklassefahrzeugen. Durch diese Entwicklung haben Verfahren, die mehr als ein Mikrofon verwenden [7, 20], an Bedeutung gewonnen.

Die vorliegende Arbeit zeigt neue, mehrkanalige Methoden zur Reduktion der mit den Mikrofonen einer Freisprecheinrichtung aufgenommenen Hintergrundgeräusche.

[1] Darunter versteht man die Anordnung mehrerer Sensoren, in diesem Fall von Mikrofonen, im Raum.

1.2 Begriffsdefinitionen

Bei der Realisierung von Freisprecheinrichtungen treten zwei Probleme auf, die aus Abbildung 1.1 ersichtlich sind: Die sich im Fahrzeuginnenraum befindlichen Mikrofone nehmen neben dem gewünschten Sprachsignal des *lokalen Sprechers* auch das über den Lautsprecher abgestrahlte Signal des *fernen Sprechers* sowie sämtliche Hintergrundgeräusche, die hier als *Fahrzeuggeräusch* bezeichnet werden, auf. Das Einkoppeln des Lautsprechersignals in die Mikrofone äußert sich als Echo auf der Seite des fernen Sprechers. Dieser hört sich selbst verzögert, was mit zunehmender Größe dieser Verzögerung den Gesprächskomfort des fernen Sprechers immer weiter reduziert. Methoden zur Unterdrückung dieser Echos werden unter dem Begriff *Echokompensation* geführt. Das mit den Mikrofonen aufgezeichnete Fahrzeuggeräusch reduziert die Sprachqualität auf der Seite des fernen Sprechers und wirkt sich somit ebenfalls störend auf die Konversation aus. Verfahren zur Reduktion dieses Hintergrundgeräuschs werden unter dem Begriff *Geräuschreduktion* zusammengefasst.

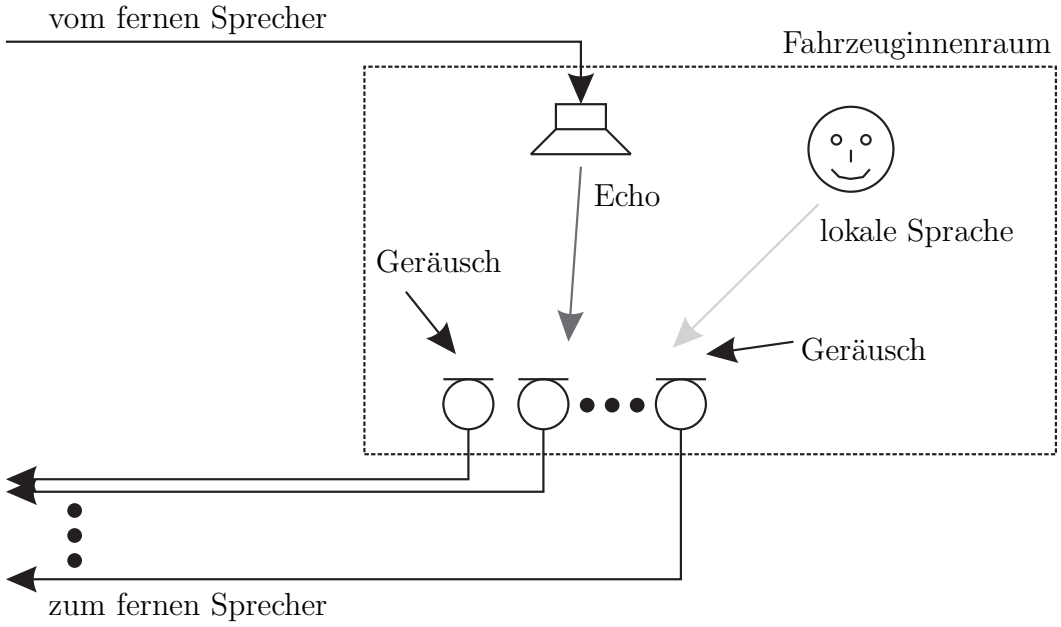

Abbildung 1.1: *Schematische Darstellung einer Freisprecheinrichtung. Das im Fahrzeuginnenraum gemessene Signal setzt sich zusammen aus: lokaler Sprache, Echo und Hintergrundgeräuschen.*

Beiden Verfahren gemein ist, dass sie jeweils dem fernen Sprecher zugute kommen, obwohl sie auf der Seite des lokalen Sprechers installiert sind. Das heißt, der Besitzer der Freisprecheinrichtung erhält keinen direkten Vorteil aus der Benutzung selbiger, sondern erhöht dadurch nur den Gesprächskomfort des fernen Sprechers, also seines Gesprächspartners. Dies ist ein Grund für die anfänglich

schlechte Akzeptanz solcher Systeme [18].

Ein vereinfachtes Signalflussdiagramm einer Freisprecheinrichtung ist in Abbildung 1.2 dargestellt. Der Block Echokompensation ist parallel zu dem aus Lautsprecher, Raum und Mikrofon (LRM) bestehenden System geschaltet. Dort wird das entstehende Echo geschätzt und vom Mikrofonsignal subtrahiert. Der Block Geräuschreduktion und Restechounterdrückung liegt dagegen im Signalpfad zum fernen Sprecher. Dort wird neben der Reduktion des Fahrzeuggeräuschs auch eine Unterdrückung des nach der Echokompensation verbliebenen Restechos durchgeführt.

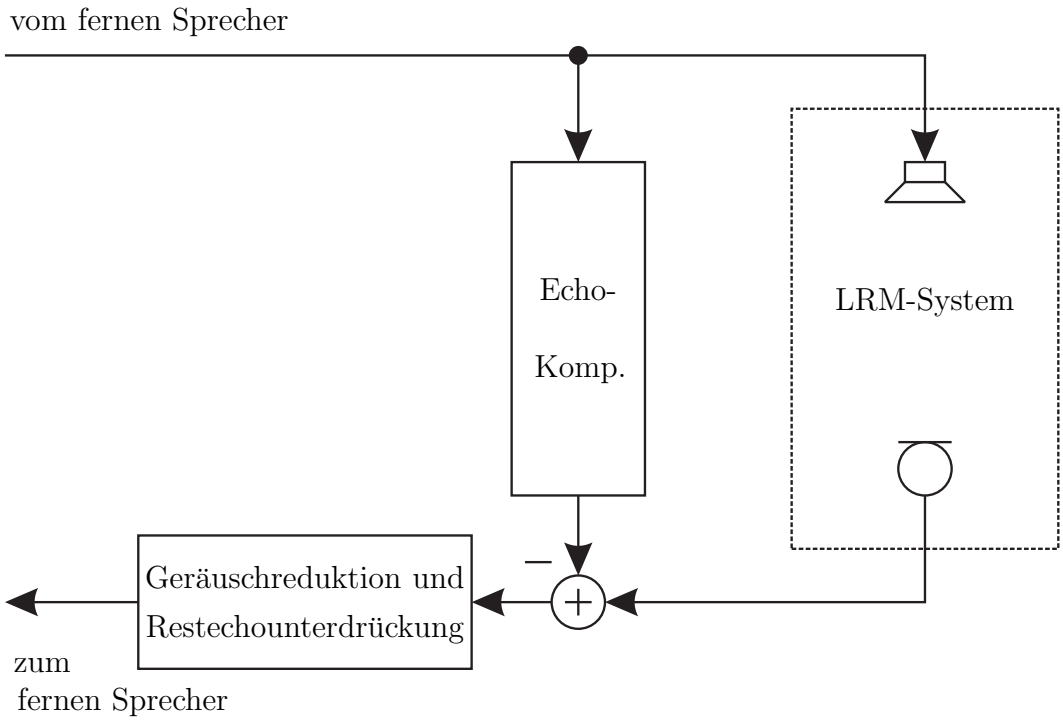

Abbildung 1.2: *Signalflussdiagramm einer Freisprecheinrichtung. Die Echokompensation ist parallel zum Lautsprecher-Raum-Mikrofon (LRM) System geschaltet, Geräuschreduktion und Restechounterdrückung seriell.*

Im mehrkanaligen Fall werden Geräusche üblicherweise in zwei Stufen reduziert [7, 18]: Zunächst werden die Mikrofonsignale $y_1(k), \ldots, y_N(k)$ einem *Beamformer* zugeführt. Darunter versteht man in der Sprachsignalverarbeitung eine Rechenvorschrift, die es ermöglicht, durch Kombination der anliegenden Mikrofonsignale alle Signalkomponenten zu unterdrücken, die nicht aus einer bestimmten Richtung des Raums, der sogenannten *Nutzrichtung*, auf das Array treffen. Bei einer Freisprecheinrichtung wird die Nutzrichtung so gewählt, dass das Array auf den Sprecher ausgerichtet ist. Die Auswirkung dieser räumlichen Filterung ent-

spricht der eines Richtmikrofons. Am einkanaligen Ausgang des Beamformers liegt das Signal $\breve{y}(k)$ an, in dem theoretisch alle Richtungen bis auf die Nutzrichtung bedämpft sind. Dadurch wird eine erste Geräuschreduktion erreicht.

In der zweiten Stufe wird dieses Ausgangssignal $\breve{y}(k)$ nun einem einkanaligen Geräuschreduktionsverfahren zugeführt. Da dieses dem Beamformer nachgeschaltet ist, wird das Filter allgemein als *Postfilter* bezeichnet. An dessen Ausgang liegt ein Schätzwert $\hat{s}(k)$ des ungestörten Sprachsignals an. Eine solche zweistufige Anordnung zur Geräuschreduktion ist für N Kanäle in Abbildung 1.3 dargestellt.

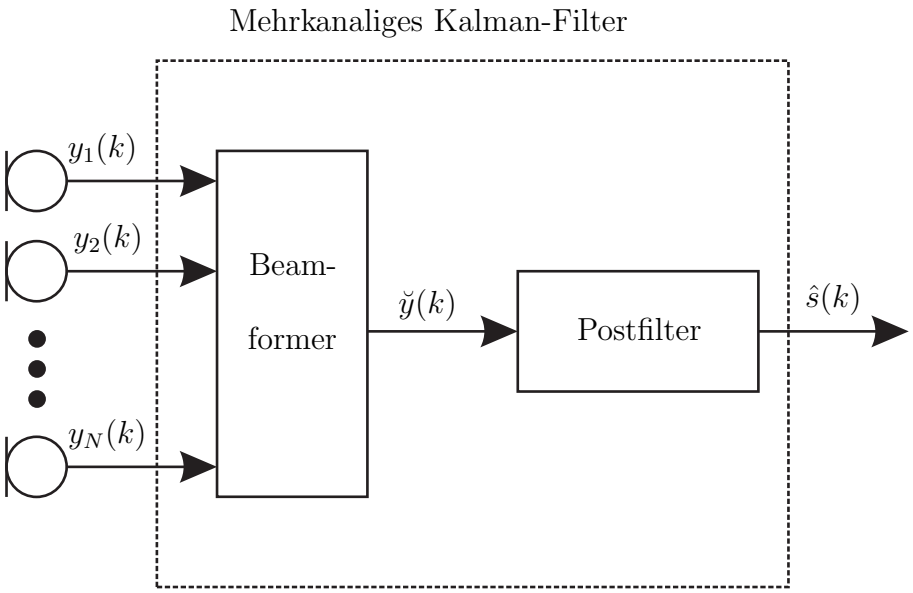

Abbildung 1.3: *Signalflussdiagramm der Beamformer-Postfilter-Struktur und deren Ersetzung durch ein mehrkanaliges Kalman-Filter (gestrichelt).*

1.3 Stand der Technik

Die Mehrzahl der heute in Freisprecheinrichtungen verwendeten einkanaligen Verfahren zur Geräuschreduktion basieren auf dem Wiener-Filter [18, 20, 52, 53], während die auf Kalman-Filterung basierten Algorithmen [16, 30, 33, 39, 58, 59] eine Minderheit darstellen. Für den Betrieb des Wiener-Filters müssen die Spektren der Sprach- und Geräuschkomponente geschätzt werden, wohingegen das Kalman-Filter die Parameter der verwendeten Modelle für Sprache und Geräusch benötigt.

In der in Abbildung 1.3 dargestellten Anordnung basiert die Mehrzahl der im Postfilter Verwendung findenden Methoden ebenfalls auf dem Konzept des Wiener-Filters [4, 7, 18]. Zudem existiert das Problem, dass Beamformer und Postfilter in vielen Systemen unabhängig voneinander operieren [7], so dass aus Sicht des Gesamtsystems die maximal mögliche Leistung nicht erreicht wird.

Darüber hinaus gibt es Verfahren, die ohne einen vorgeschalteten Beamformer direkt den N-kanaligen Eingang in einem mehrkanaligen Wiener-Filter verarbeiten [5]. Für diesen Anwendungsfall sind keine auf Kalman-Filterung basierten Verfahren bekannt. Lediglich das in [2] vorgeschlagene Kalman-Filter verwendet zwei Mikrofone, wobei eines aber als Referenz dient, was nicht dem klassischen Ansatz eines Beamformers entspricht.

1.4 Ziele dieser Arbeit

Zielsetzung dieser Arbeit ist die Erweiterung des in [39] beschriebenen einkanaligen Kalman-Filter-Verfahrens auf den mehrkanaligen Fall. Hierbei werden die nachstehenden Ziele verfolgt:

- Optimierung des Gesamtsystems bestehend aus Beamformer und Postfilter durch Entwurf einer mehrkanaligen Kalman-Filter-Struktur, die die Funktionalität beider Teilsysteme kombiniert (siehe Abbildung 1.3).
- Entwurf mehrkanaliger Verfahren zur verbesserten Schätzung der für das Kalman-Filter benötigten Parameter.
- Erhöhung der Effizienz durch Wahl einer hinsichtlich Aufwand und Flexibilität optimierten Implementierungsstruktur.

Der Schwerpunkt liegt dabei stets auf dem durch die Mehrkanaligkeit erzielbaren Gewinn. Aspekte der Echokompensation sowie der Restechounterdrückung bleiben im Folgenden unberücksichtigt.

Ausgehend vom einkanaligen Kalman-Filteransatz werden durch die Hinzunahme weiterer Mikrofone an folgenden Stellen Verbesserungen erwartet:

- Bei der Schätzung der Sprachmodellparameter, da jetzt das von N Mikrofonen aufgenommene verrauschte Sprachsignal des lokalen Sprechers vorliegt.
- Durch die Verwendung eines mehrkanaligen Filter-Algorithmus an sich, da das Ausgangssignal $\hat{s}(k)$ jetzt aus mehr als einem Eingang berechnet wird.

1.5 Gliederung der Arbeit

Die vorliegende Arbeit ist wie folgt gegliedert: Nach dieser *Einleitung* werden im zweiten Kapitel *Signalanalyse und -modellierung* die in einer Freisprecheinrichtung vorkommenden Signale, das sind Sprache und Fahrzeuggeräusch, analysiert. Darauf aufbauend wird anschließend die Modellierung von Sprache behandelt sowie die für diese Arbeit verwendeten Audiodaten beschrieben.

Das dritte Kapitel *Signalmodell und Kalman-Filter* beschäftigt sich mit Themen, die den Kalman-Filter-Algorithmus betreffen. Dazu wird zunächst das zugrunde liegende Signalmodell vorgestellt. Anschließend wird dieses verwendet, um ein darauf angepasstes mehrkanaliges Kalman-Filter herzuleiten. Der Rest des Kapitels befasst sich mit Diskussion und Interpretation der Eigenschaften des gefundenen Algorithmus'. Dies beinhaltet die Beschreibung von Methoden zur Verbesserung der Robustheit sowie die Betrachtung des numerischen Aufwands.

Im vierten Kapitel *Schätzung der Filterparameter* wird die Schätzung der für die Benutzung des Kalman-Filters notwendigen Parameter behandelt. Der Schwerpunkt liegt dort auf der Schätzung der Sprachmodellparameter. Begonnen wird das Kapitel mit einer theoretischen Betrachtung der verwendeten autoregressiven Signalmodellierung. Dabei wird insbesondere der Zusammenhang mit dem linearen Prädiktor aufgezeigt. Daran anschließend werden verschiedene Methoden der parametrischen und nicht-parametrischen Spektralschätzung vorgestellt, soweit sie für die vorgeschlagenen Schätzverfahren benötigt werden. Diese mehrkanaligen Verfahren werden im dritten Teil des Kapitels beschrieben und bezüglich ihres numerischen Aufwands verglichen. Zum Schluss wird die Schätzung der den Fahrzeuginnenraum beschreibenden Größen vorgestellt.

Das fünfte Kapitel *Implementierung des Gesamtsystems* behandelt zuerst die verwendete Filterbankstruktur und zeigt deren Vor- und Nachteile gegenüber anderen Verarbeitungsstrukturen, wie zum Beispiel der Vollbandverarbeitung, auf. Anschließend wird auf die Parametrierung der zeitinvarianten Parameter eingegangen. Darunter fallen insbesondere die in den einzelnen Teilbändern verwendeten Sprachmodellordnungen, deren Festlegung dargelegt wird. Das Kapitel schließt mit der Betrachtung der durch das System verursachten Signalverzögerung und den sich daraus ergebenden Möglichkeiten der Nachglättung.

Im sechsten Kapitel *Bewertung der Simulationsergebnisse* werden Simulationsergebnisse vorgestellt und diskutiert sowie ein abschließendes Fazit gezogen. Das Hauptaugenmerk liegt dabei sowohl auf der erzielbaren Güte der Schätzung der Sprachmodellparameter als auch auf der erreichten Geräuschreduktion des Gesamtsystems. Abgeschlossen wird diese Arbeit mit dem siebten Kapitel *Zusammenfassung und Ausblick*.

Kapitel 2

Signalanalyse und -modellierung

Das mit dem Mikrofon einer Freisprecheinrichtung in einem Kraftfahrzeug aufgenommene Signal besteht aus zwei Hauptkomponenten [4, 19, 39]:

- Sprache und
- Fahrzeuggeräusch.

In dieser Arbeit wird angenommen, dass sich beide Komponenten additiv überlagern. Ziel aller Verfahren zur Geräuschreduktion ist es, den Geräuschanteil im Summensignal so weit wie möglich zu reduzieren, und dabei gleichzeitig die Sprachkomponente mit minimaler Verzerrung zu erhalten.

Wie sich im Verlauf dieses Kapitels zeigen wird, sind die Voraussetzungen dafür ungünstig. Zum einen nehmen Sprache und Fahrzeuggeräusch bis auf den Frequenzbereich unterhalb der Sprachgrundfrequenz (siehe Abschnitt 2.1.1), in dem nur Geräusch vorliegt, etwa den gleichen Frequenzbereich ein. Zum anderen ähneln sich ihre spektralen Verläufe mit kleiner werdender Signalleistung bei größer werdender Frequenz.

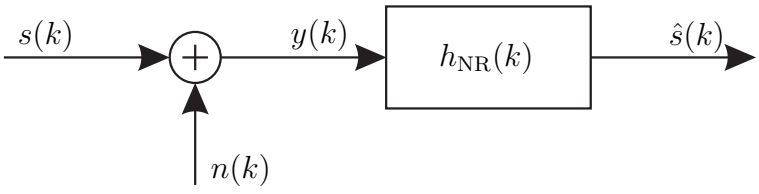

Abbildung 2.1: *Allgemeines Blockschaltbild des Geräuschreduktionsproblems.*

Das Grundschema der Geräuschreduktion ist in Abbildung 2.1 als Blockschaltdiagramm abgebildet, wobei $s(k)$ das Sprach-, $n(k)$ das Geräusch- und $y(k)$ das Summensignal bezeichnet. Außerdem wurde die eigentliche Geräuschreduktion symbolisch als Filter mit der Impulsantwort $h_{\mathrm{NR}}(k)$ dargestellt, dessen Ausgang

eine Schätzung[1] $\hat{s}(k)$ der Sprachkomponente $s(k)$ ist.

In diesem Kapitel werden die oben genannten Hauptkomponenten analysiert, ihre Eigenschaften diskutiert und das im später beschriebenen Verfahren verwendete Signalmodell, welches für die Herleitung des Kalman-Filters benötigt wird, abgeleitet. Im letzten Teil werden die zur Simulation verwendeten Audiodaten beschrieben. Die in den nachfolgenden Kapiteln vorgestellten Methoden basieren größtenteils auf Annahmen, die aus Erkenntnissen dieses Kapitels resultieren.

2.1 Analyse von Sprache

Dieser Abschnitt teilt sich in drei Teile. Zunächst werden Sprachsignale im Zeitbereich analysiert. Danach werden im Frequenzbereich die spektralen Eigenschaften diskutiert. Abschließend wird kurz die Erzeugung von Sprache im menschlichen Körper behandelt.

2.1.1 Zeitbereichsanalyse

In Abbildung 2.2 ist ein Ausschnitt eines Sprachsignals im Zeitbereich dargestellt. Anhand dieses etwa zweieinhalb Sekunden langen Beispiels lässt sich bereits ein signifikantes Merkmal von Sprache erkennen.

Auffallend ist, dass Abschnitte großer Amplitude und Signalleistung (die sogenannten *Sprachphasen*) sich mit solchen abwechseln, in denen nahezu keine Signalleistung vorhanden ist (die sogenannten *Sprachpausen*). Betrachtet man nur die Sprachphasen, so fallen auch hier starke Amplitudenschwankungen auf. Dieses hochgradig zeitveränderliche Verhalten stellt eine der Hauptschwierigkeiten bei der Signalverarbeitung von Sprachsignalen dar. Zudem müssen bei vielen Geräuschreduktionsverfahren die Übergänge von Sprachpausen zu Sprachphasen und umgekehrt im verrauschten Summensignal detektiert werden[2].

Im später vorgestellten Kalman-Filter werden alle vorkommenden Signale als stochastische Prozesse modelliert. Diese werden durch ihre Momente (z.B. linearer Mittelwert, Autokorrelationsfunktion, etc.) beschrieben, die über die Schar der Realisierungen, den sogenannten Musterfunktionen, gebildet werden. In der Realität, beispielsweise bei der Aufnahme mit einem Mikrofon, wird jeweils nur eine dieser Musterfunktionen gemessen. Daher können nur Zeitmittelwerte berech-

[1] Der Dachoperator (ˆ) bezeichnet in dieser Arbeit grundsätzlich einen Schätzwert einer Größe. Beispielsweise bezeichnet $\hat{x}(k)$ einen Schätzwert für $x(k)$.

[2] Diese unter dem Begriff *Sprachpausendetektion* bzw. im Englischen *Voice Activity Detection (VAD)* zusammengefassten Verfahren werden in Kapitel 4 kurz erläutert.

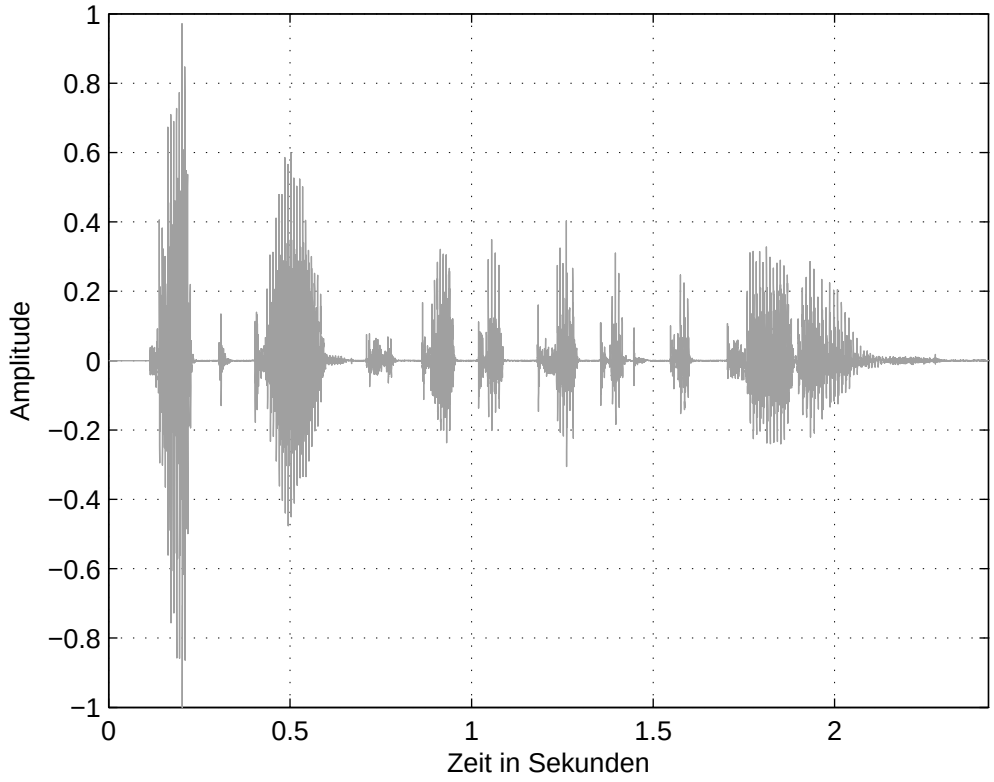

Abbildung 2.2: *Ausschnitt eines Sprachsignals eines Mannes.*

net werden. Damit diese aussagekräftig für die Scharmittelwerte sind, muss die gemessene Musterfunktion repräsentativ für den Zufallsprozess sein. Sind mit Wahrscheinlichkeit Eins alle Musterfunktionen repräsentativ, so nennt man den Prozess *ergodisch* [19]. Ergodizität setzt *Stationarität* voraus, wobei zwischen strenger und schwacher Stationarität unterschieden wird. Während für strenge Stationarität alle Momente zeitinvariant sein müssen, liegt schwache Stationarität bereits vor, wenn die Momente bis zur zweiten Ordnung – das sind linearer und quadratischer Mittelwert sowie die Autokorrelationsfunktion – zeitinvariant sind.

Aufgrund des stark zeitveränderlichen Verhaltens kann für Sprachsignale im Allgemeinen keine Stationarität (auch keine schwache) angenommen werden. Man behilft sich durch blockweise Verarbeitung des Datenstroms, wobei die Blocklänge so klein gewählt wird, dass das Signal innerhalb dieser Blöcke als schwach stationär modelliert werden kann. In diesem Fall spricht man von Kurzzeit- bzw. Quasistationarität. Diese Länge kann bei Sprachsignalen zwischen 20 und 400 ms variieren. Um möglichst immer Kurzzeitstationarität gewährleisten zu können,

hat sich daher in der Sprachsignalverarbeitung eine Blocklänge von 20 bis 50 ms durchgesetzt [52].

2.1.2 Frequenzbereichsanalyse

Für die Analyse des Spektralverhaltens von Sprache wird die Kurzzeit-Fouriertransformation gewählt:

$$S(e^{j\Omega}, \eta) \;=\; \sum_{k=0}^{L-1} s_\eta(k) e^{-j\Omega k}. \tag{2.1}$$

Im Gegensatz zu der zeitdiskreten Fouriertransformation, bei der über alle Grenzen summiert wird, werden die Kurzzeitspektren $S(e^{j\Omega}, \eta)$ aus Blöcken $s_\eta(k)$ der Länge L des Eingangssignals $s(k)$ berechnet. Diese werden durch den Blockindex η unterschieden. Ohne Einschränkung der Allgemeinheit können aufeinander folgende Blöcke unmittelbar aneinander anschließen oder sich überlappen. Für die Zerlegung des Eingangssignals $s(k)$ in Blöcke wird neben der Blocklänge L auch der Blockversatz Q, das ist der in Abtastwerten gemessene Abstand von zwei aufeinander folgenden Blöcken, angegeben. Die Anzahl der überlappenden Abtastwerte ergibt sich folglich zu $L-Q$. Für den η-ten Block gilt demnach:

$$s_\eta(k) = s(k - \eta Q) \quad \text{für} \quad k = 0, \ldots, L-1. \tag{2.2}$$

Eine alternative Darstellung der blockweisen Datenverarbeitung ist die Verwendung von Vektoren, welche bei der Herleitung des Kalman-Filters in Kapitel 3 erklärt und benutzt wird.

Bestimmt man von aufeinander folgenden (ggf. auch überlappenden) Blöcken die Kurzzeitspektren und ordnet diese in einem Zeit-Frequenz-Diagramm an, so erhält man eine weitere Möglichkeit der Visualisierung: das *Spektrogramm* [18]. Dabei werden die spektralen Amplituden durch verschiedene Graustufen, von weiß für sehr kleine bis schwarz für sehr große Amplituden, symbolisiert. Diese frequenzabhängige Spektraldarstellung eignet sich besonders gut, um Signale mit ausgeprägtem zeitveränderlichem Verhalten zu analysieren. In Abbildung 2.3 ist das Spektrogramm des in Abbildung 2.2 als Zeitbereichssignal dargestellten Sprachsignals abgebildet. Dafür wurden bei $f_s = 8$ kHz Abtastfrequenz Blöcke der Länge 512 Abtastwerte (Frequenzauflösung) und einer Überlappung aufeinander folgender Blöcke von 500 Abtastwerten (Zeitauflösung) gewählt. Die Zeitauflösung hängt zudem indirekt von der Blocklänge ab, da zeitliche Veränderungen, die deutlich kürzer als die Blocklänge sind, nicht mehr dargestellt werden können.

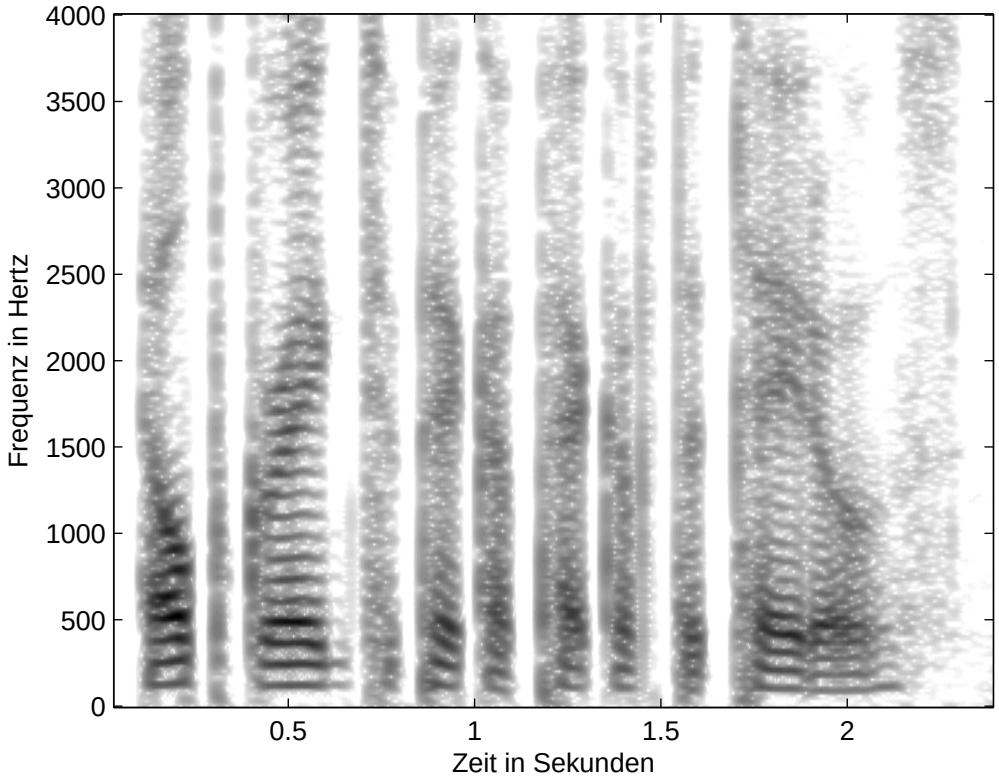

Abbildung 2.3: *Spektrogramm eines Sprachsignals eines Mannes.*

Neben dem auch hier sichtbaren zeitveränderlichen Verhalten, fällt besonders die abschnittsweise harmonische Struktur auf (beispielsweise bei $t = 0,5$ s). Darüber hinaus enthält das Spektrogramm auch Abschnitte, die Rauschen ähneln (zum Beispiel bei $t = 0,3$ s). Diese verschiedenen Ausprägungen von Sprache werden im nächsten Abschnitt behandelt.

2.1.3 Erzeugung von Sprache im menschlichen Körper

Sprache wird beim Menschen in den Sprechorganen erzeugt. Dies sind die Lungen als Energielieferant, die Luftröhre, der Kehlkopf mit den Stimmbändern (Glottis) sowie der Mund-, Rachen- und Nasenraum. Letztere werden zusammengefasst als *Vokaltrakt*[3] bezeichnet. Die so erzeugten Schallwellen werden über die Lippen und den Mund abgestrahlt.

[3]Das Wort leitet sich nicht von den Vokalen ab, sondern von dem lateinischen Wort *vox* für Stimme [52].

Die Aufgaben der einzelnen Organe können in zwei Hauptgruppen unterteilt werden: Anregung und Signalformung. Während Lunge, Kehlkopf und Stimmbänder die Anregung der verschiedenen Laute übernehmen, wird deren Ausformung im Vokaltrakt vorgenommen. Dabei werden drei Arten von Anregung unterschieden:

- stimmhafte Anregung,
- stimmlose Anregung
- und transiente Anregung.

Stimmhafte Anregung

Die im letzten Abschnitt erwähnte harmonische Struktur entsteht bei der stimmhaften Anregung, die beispielsweise bei der Artikulation der Vokale ([a], [e], [i], etc.) vorliegt. Dabei schwingen die Stimmbänder mit der als *Sprachgrundfrequenz* bzw. unter Verwendung des englischen Begriffs als *Pitchfrequenz* bezeichneten Frequenz und deren Oberwellen. Sie variiert etwa zwischen 50 Hz bei tiefen Männerstimmen und 300 bis 400 Hz bei hohen Frauenstimmen. Für die deutsche Sprache können die folgenden Mittelwerte angenommen werden [1]: 120 Hz für Männer und 215 Hz für Frauen. Zudem kann jeder Sprecher seine Pitchfrequenz in gewissen Grenzen variieren, wodurch sich eine Satzmelodie ausprägen kann [39].

Aufgrund der Beschaffenheit der Stimmbänder ist besagte harmonische Struktur nur in den seltensten Fällen exakt [23]. Das heißt, die Oberwellen liegen nur ungefähr bei ganzzahligen Vielfachen der Pitchfrequenz, wobei die Abweichung von der idealen harmonischen Struktur mit steigender Frequenz größer wird. Daher wird in diesem Zusammenhang von einer *quasi-harmonischen* Struktur gesprochen. Dies muss beispielsweise bei Geräuschreduktionsverfahren, die auf der Schätzung der Sprachgrundfrequenz und anschließender Filterung mittels Kammfilter beruhen, beachtet werden. Hier muss die Filterdämpfung mit zunehmender Frequenz reduziert werden, um zu verhindern, dass bei höheren Frequenzen aufgrund der nicht exakt harmonischen Struktur und des Schätzfehlers bei der Bestimmung der Sprachgrundfrequenz Oberschwingungen unterdrückt anstatt durchgelassen werden [48, 49].

Das in Abbildung 2.4 dargestellte Kurzzeitspektrum zeigt solch einen stimmhaften Laut. Wie zu erkennen ist, liegt die Pitchfrequenz ziemlich genau bei $f_p = 100$ Hz. Darüber hinaus kann beobachtet werden, dass mit steigender Frequenz, wie soeben beschrieben, die Vielfachen der Sprachgrundfrequenz nicht mehr ganz genau getroffen werden. Neben der relativ schnellen Schwingung der Pitchfrequenz, kann noch eine deutlich langsamere erkannt werden. Diese als Formanten bezeichneten lokalen Maxima werden bei der Diskussion des Sprach-

modells in Abschnitt 2.3 genauer erläutert.

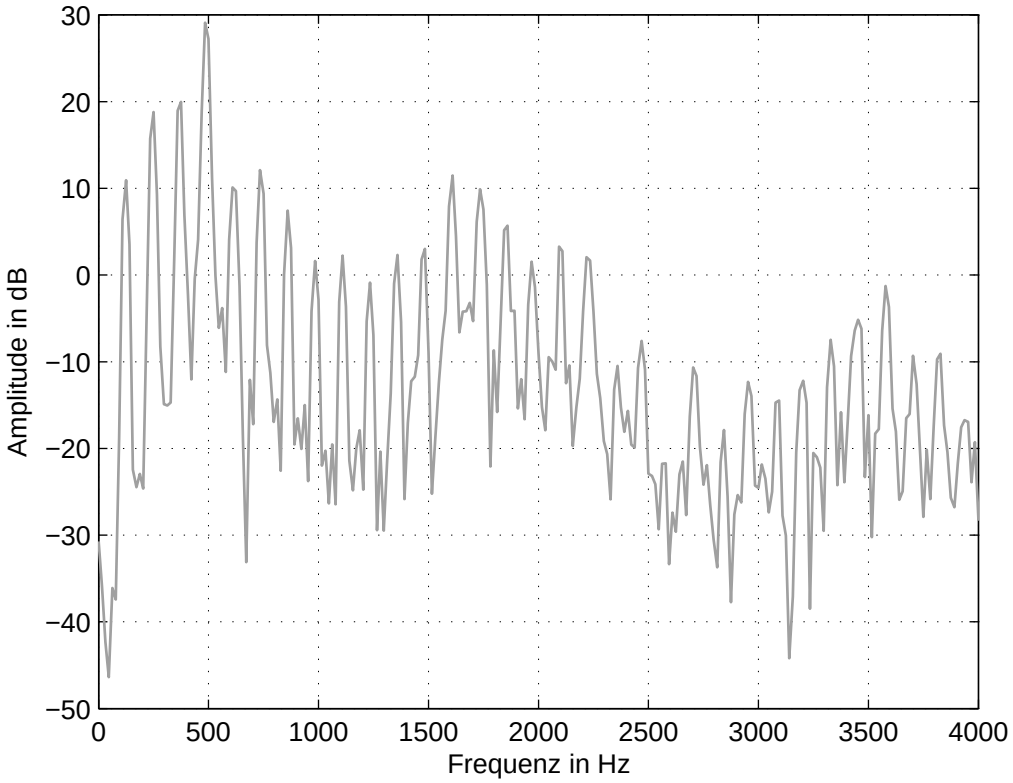

Abbildung 2.4: *Kurzzeit-Spektrum eines stimmhaften Lautes.*

Stimmlose Anregung

Bei der stimmlosen Anregung liegt keine Schwingung der Stimmbänder vor. Die aus der Lunge strömende Luft wird an den Verengungen des Vokaltrakts verwirbelt, so dass eine turbulente Strömung entsteht. Das resultierende Signal hat daher eine spektrale Einhüllende, die der von Rauschsignalen ähnelt. Stimmlose Anregung wird für Zischlaute wie das [f] oder [s] verwendet. Der rauschartige Charakter ist gut in Abbildung 2.5 zu erkennen, die das Kurzzeitspektrum eines stimmlosen Lautes zeigt. Stimmhafte und stimmlose Anregung können auch zeitgleich vorliegen. In diesem Fall wird von gemischter Anregung gesprochen.

Transiente Anregung

Die transiente Anregung liegt bei der Artikulation von Plosivlauten, wie z.B. dem [p] oder [t] vor. Dabei wird der Luftstrom nach Passieren der Stimmbänder (üblicherweise an den Lippen) kurzzeitig unterbrochen und danach schlagartig (explosionsartig) wieder freigegeben [28].

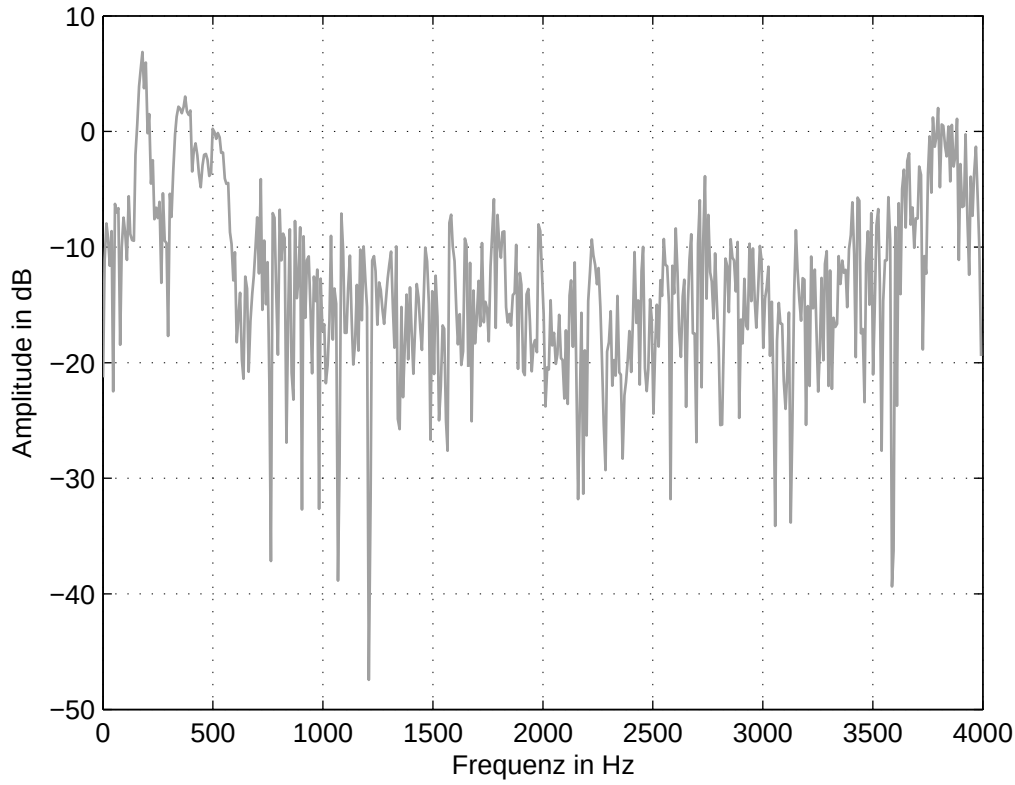

Abbildung 2.5: *Kurzzeit-Spektrum eines stimmlosen Lautes.*

2.2 Analyse von Fahrzeuggeräusch

Das als Fahrzeuggeräusch bezeichnete Hintergrundrauschen im Innenraum eines fahrenden Kraftfahrzeuges setzt sich aus drei Hauptanteilen zusammen, die im Folgenden kurz erläutert werden [39]:

- Motorgeräusch,
- Reifen- bzw. Rollgeräusch
- und Windgeräusch.

Daneben haben auch andere mechanisch bewegte Teile, wie z.B. das Getriebe oder Vibrationen der Karosserie, Anteile im Gesamtgeräusch. Da diese Komponenten, genauso wie transiente Störungen, die beispielsweise beim Überfahren von Löchern, Schwellen, etc. entstehen, hauptsächlich nur zeitweise auftretenden, werden sie bei dieser Analyse vernachlässigt.

2.2.1 Motorgeräusch

Das Motorgeräusch stellt bei einigen Kraftfahrzeugen – vor allem bei sogenannten Sportwagen – die dominante Komponente im Gesamtgeräusch dar. Es ist charakterisiert durch seine harmonische Struktur, die durch die Gas- und Massenkräfte verursacht wird und deren genaue Ausprägung von der Zylinderanzahl sowie der Art des Motors (4-Takt-, 2-Takt- oder Dieselmotor) abhängig ist. Zum Beispiel weisen die Spektren von 4-Takt- und Dieselmotoren Vielfache der halben Motordrehzahl auf. Dies rührt daher, dass bei dieser Motorenauslegung pro Zylinder eine Zündung alle zwei Umdrehungen erfolgt. Je nachdem wie viele Zylinder solch ein Motor aufweist, sind manche Vielfache besonders stark ausgeprägt. Bei einer 4-Zylinderanordnung sind dies beispielsweise die dritte und fünfte Oberwelle [15]. Allgemein kann gesagt werden, dass die Ausprägung des Motorgeräuschs umso stärker ist, je höher die Drehzahl und das dem Motor abverlangte Drehmoment ist. Darüber hinaus ist sein Einfluss auf das Gesamtgeräusch für Frequenzen oberhalb etwa 1000 Hz nicht mehr signifikant. Abhängig von der Situation (z.B. Stadtverkehr oder Autobahn) und dem Fahrstil des jeweiligen Fahrers, kann das Motorgeräusch schnell veränderlich oder auch relativ konstant sein.

2.2.2 Reifen- bzw. Rollgeräusch

Das Reifen- bzw. Rollgeräusch entsteht durch das Abrollen der Reifen auf dem jeweiligen Untergrund. Es ist abhängig von der Beschaffenheit des Untergrunds und der Reifen, sowie von der Geschwindigkeit. Seine logarithmierte Leistung wächst in erster Näherung linear mit steigender Geschwindigkeit [39]. Die spektrale Einhüllende kann als tiefpassgefärbtes Rauschen charakterisiert werden. Seine statistischen Eigenschaften ändern sich im Allgemeinen nur langsam über der Zeit, so dass es näherungsweise als stationär modelliert werden kann.

2.2.3 Windgeräusch

Das Windgeräusch entsteht durch die abrupte Ablenkung und damit Verwirbelung des Luftstroms hinter Kanten wie zum Beispiel den Außenspiegeln, Radkästen und Scheibenwischern. Es ist nur von der Geschwindigkeit abhängig und trägt außer bei hohen Geschwindigkeiten nur gering zum Gesamtgeräusch bei.

2.2.4 Zusammenfassung und Vergleich mit Sprache

Zusammenfassend kann gesagt werden, dass Fahrzeuggeräusch einen tiefpassgefärbten, rauschartigen Charakter aufweist. Dabei tritt die harmonische Struktur des Motorgeräuschs im Gesamtsignal nur bei starken Beschleunigungen und nur im tiefen Frequenzbereich in Erscheinung. Somit können seine statistischen Eigenschaften im Allgemeinen als langsam veränderlich angesehen werden. Ein typisches Spektrogramm von Fahrzeuggeräusch bei einer Fahrt auf der Autobahn mit einer konstanten Geschwindigkeit von ca. 160 km/h ist in Abbildung 2.6 dargestellt.

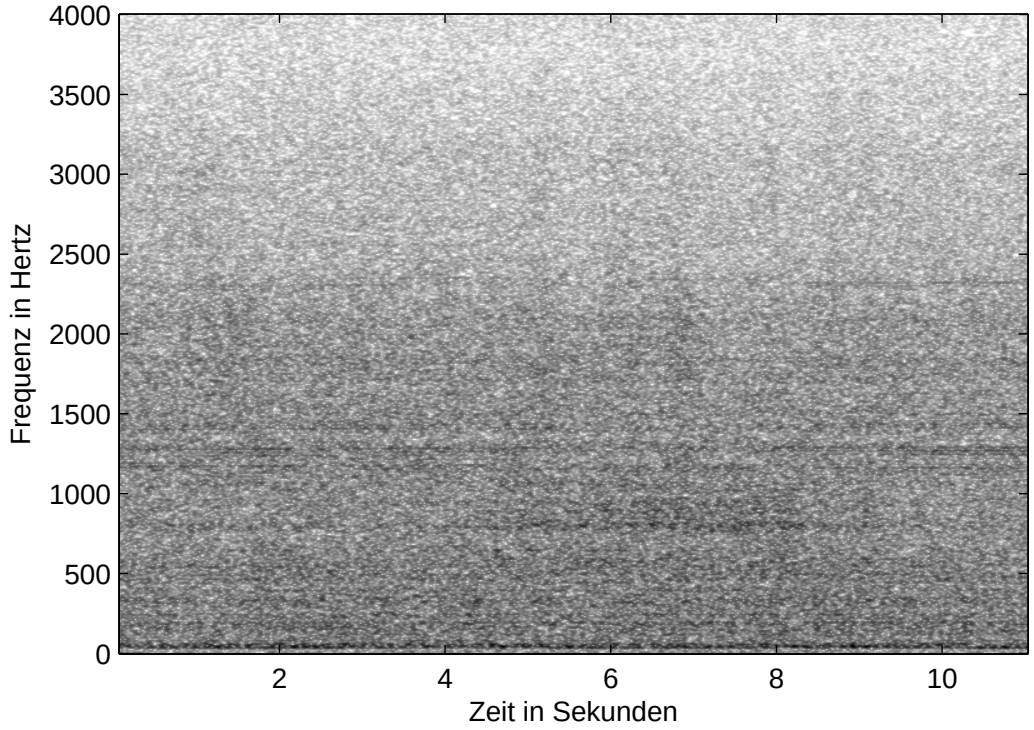

Abbildung 2.6: *Spektrogramm von Fahrzeuggeräusch bei einer Geschwindigkeit von etwa 160 km/h.*

Am Anfang des Kapitels wurde bereits erwähnt, dass Sprache und Fahrzeuggeräusch den gleichen Frequenzbereich einnehmen, was eine klassische Filterung zu deren wirkungsvoller Trennung ausschließt. In Abbildung 2.7 sind gemittelte Kurzzeitspektren von Sprache und Fahrzeuggeräusch abgebildet. Die Ähnlichkeit des spektralen Verlaufs ist offensichtlich.

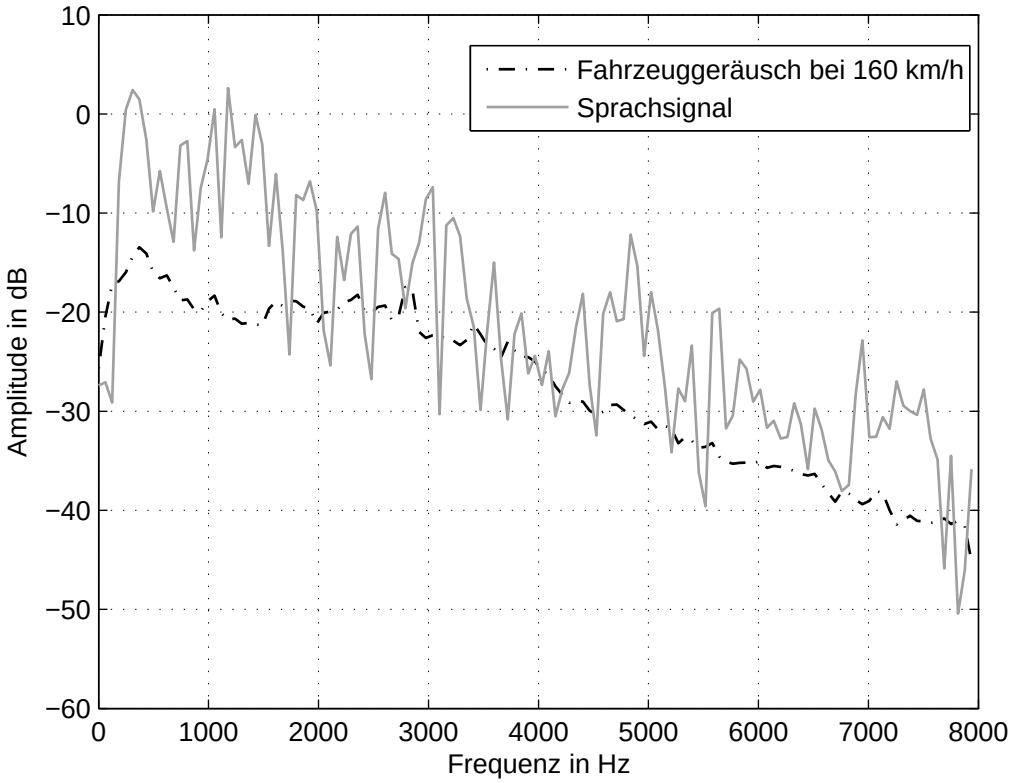

Abbildung 2.7: *Gemittelte Kurzzeitspektren von Sprache und Fahrzeuggeräusch.*

2.3 Modellierung von Sprache

Im Folgenden wird das in dieser Arbeit verwendete Quelle-Filter-Modell, welches sich zur Modellierung von Sprache als leistungsfähig erwiesen hat [45, 53, 28], vorgestellt. Bei dieser Modellierung werden die oben beschriebenen Sprechorgane mathematisch nachgebildet, wobei ebenfalls die Unterteilung in Anregung (Quelle) und Signalformung (Filter) übernommen wird. Das Blockschaltbild dieses Modells ist in Abbildung 2.8 dargestellt.

Die stimmhafte Anregung wird entsprechend den physikalischen Gegebenheiten

im menschlichen Körper durch einen Impulsgenerator (IG) mit einem nachgeschalteten Filter $H_{\text{glot}}(z)$, welches die Stimmbänder (Glottis) modelliert, abgebildet. Im Gegensatz dazu wird für die stimmlose Anregung nur ein Rauschgenerator (RG) ohne nachgeschaltetes Filter benutzt. Dieser zweite Teil der Anregungsmodellierung entspricht nicht den physikalischen Gegebenheiten, da das Rauschen in Wirklichkeit im Vokaltrakt erzeugt wird.

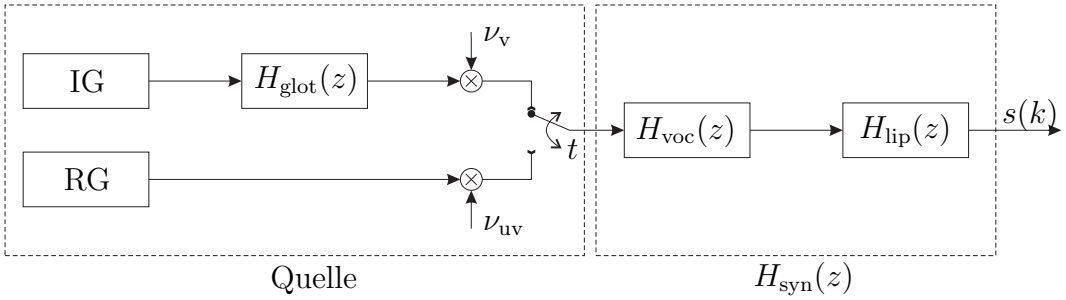

Abbildung 2.8: *Quelle-Filter-Modell zur Spracherzeugung und dessen Vereinfachung (gestrichelt).*

Die Amplituden dieser beiden Quellensignale werden jeweils mit einem spezifischen Verstärkungsfaktor (ν_{v} bzw. ν_{uv}) angepasst. Anschließend sorgt der Schalter t dafür, dass entweder das Signal der stimmhaften oder das Signal der stimmlosen Anregung weitergeleitet wird. Da keine Möglichkeit besteht, beide Signale gleichzeitig durchzuschalten, kann die gemischte Anregung mit diesem Ansatz nicht modelliert werden. Gleiches gilt für die plosive Anregung, die nicht explizit abgebildet wird.

Die beiden Anregungssignale weisen bis zu diesem Punkt neben der ggf. vorhandenen Pitchstruktur bei stimmhafter Anregung keine besondere spektrale Form auf. Sie sind spektral flach. Ihre spektrale Ausprägung wird im Vokaltraktfilter $H_{\text{voc}}(z)$ vorgenommen. Dabei werden vor allem die Formanten, die durch die Resonanzfrequenzen des Vokaltrakts entstehen, dem späteren Ausgangssignal aufgeprägt. Zuletzt wird noch die Abstrahlcharakteristik der Lippen durch das Filter $H_{\text{lip}}(z)$ nachgebildet, welches schließlich das Sprachsignal $s(k)$ ausgibt.

Das oben beschriebene Vorgehen von der Anregung bis zum fertigen Sprachsignal wird als Sprachsynthese bezeichnet. Der umgekehrte Weg – die Sprachanalyse – ist ebenfalls möglich. Dazu werden die einzelnen Filter auf ein gegebenes Sprachsignal angepasst. Führt man danach eine inverse Filterung durch, so erhält man bei idealer Anpassung das zugrunde liegende Anregungssignal. Damit ist es möglich, ein Sprachsignal zunächst in das Anregungssignal und die Formfilterparameter zu zerlegen und anschließend durch Filterung wieder vollständig zu rekonstruieren. Dieses Prinzip ist in Abbildung 2.9 dargestellt.

Bei nicht-idealer Anpassung, wie sie bei real gemessenen Signalen auftritt, ent-

spricht das bei der Analyse erhaltene Signal nicht dem Anregungssignal, sondern weist einen zusätzlichen Restfehler auf, der bei der Modellierung im Signalmodell bzw. Kalman-Filter berücksichtigt werden muss. Auch hier kann durch Filterung das ursprüngliche Signal vollständig rekonstruiert werden. Diese Zusammenhänge sind stark mit dem Verfahren der linearen Prädiktion – hier insbesondere mit dem Prädiktorfehlerfilter, welches ebenfalls ein Signal durch Prädiktorkoeffizienten und Restfehler darstellt – verwandt. Auf die lineare Prädiktion wird sowohl bei der Herleitung des Kalman-Filters in Kapitel 3 als auch bei der Parameterschätzung in Kapitel 4 eingegangen.

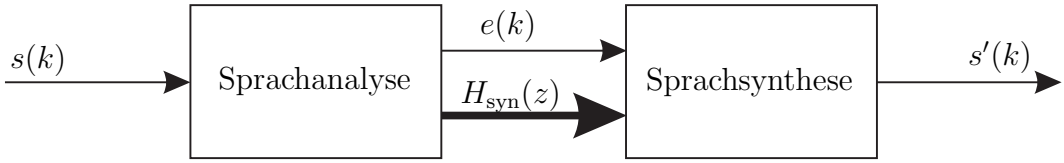

Abbildung 2.9: *Prinzip der Sprachanalyse und -synthese. Das Sprachsignal $s(k)$ wird durch Formfilterparameter $H_{syn}(z)$ und Restfehlersignal $e(k)$ dargestellt und hieraus anschließend wieder zusammengesetzt. Es gilt: $s(k) = s'(k)$ (perfekte Rekonstruktion).*

Für das in dieser Arbeit verwendete Sprachmodell wird das lineare Quelle-Filter-Modell vereinfacht. Diese Vereinfachung ist in Abbildung 2.8 bereits angedeutet. Zunächst wird die Unterscheidung zwischen stimmhaften und nicht stimmhaften Lauten fallen gelassen und durch eine einzige Quelle ersetzt. Dieses Vorgehen lässt sich damit begründen, dass das Modell in einer Analyse-Synthese-Struktur verwendet wird. Das heißt, dass zu dessen Anregung das Fehlersignal der Analysestufe benutzt wird, in dem die jeweilige Anregung bereits enthalten ist. Des Weiteren werden die Filter $H_{\text{voc}}(z)$ und $H_{\text{lip}}(z)$ in einem Formfilter $H_{\text{syn}}(z)$ zusammengefasst, um die Schätzung der Filterparameter zu vereinfachen. Dabei wird der Einfluss von $H_{\text{lip}}(z)$ nahezu vollständig vernachlässigt, da es für die Signalformung als Ganzes nur eine untergeordnete Rolle spielt.

Für eine vollständige Modellierung des Vokaltrakts, das heißt inklusive des Nasaltrakts, muss für $H_{\text{syn}}(z)$ die Form eines Pol-Nullstellen-Filters gewählt werden [53]. Dies ist gleichbedeutend mit der Verwendung der sogenannten autoregressiven moving average (ARMA) Modellierung. Vernachlässigt man dagegen den Einfluss des Nasaltrakts, kann vereinfachend auch ein rein rekursives Polstellen-Filter benutzt werden. In diesem Fall spricht man von einer autoregressiven (AR) Modellierung. Der Vorteil dieser Filterstruktur liegt in der engen Verwandschaft zur parametrischen Spektralschätzung mittels linearer Prädiktion.

Aus diesem Grund wird diese im folgenden Kapitel 3 für die Herleitung des Kalman-Filters verwendet. Die mathematische Beschreibung der oben genannten Filter wird ausführlich in Kapitel 4 behandelt.

2.4 Beschreibung der verwendeten Audiodaten

Die in dieser Arbeit verwendeten Audiodaten[4] wurden nicht direkt durch Aufnahme von Sprache in einem fahrenden Kraftfahrzeug aufgenommen, sondern auf Grundlage von gemessenen Impulsantworten und aufgezeichnetem Fahrzeuggeräusch künstlich generiert. Dieses Vorgehen hat den Vorteil, dass beliebige Sprachsequenzen in dem einmal vermessenen Kraftfahrzeug simuliert werden können, ohne dass jedesmal wieder aufwendige Messfahrten notwendig werden. Die Generierung der Audiosignale wird im Folgenden beschrieben.

Die Grundidee besteht darin, aus einem gegebenen ungestörten Sprachsignal $s(k)$, einer kausalen Raumimpulsantwort $h(k)$ der Länge L_h und Fahrzeuggeräusch $n(k)$ das Mikrofonsignal $y(k)$ zu berechnen. Dazu wird $s(k)$ mit $h(k)$ gefaltet und das Ergebnis mit $n(k)$ additiv überlagert[5]:

$$y(k) \quad = \quad s(k) * h(k) + \beta_{\mathrm{SNR}} n(k) = \sum_{l=0}^{L_h} h(l)s(k-l) + \beta_{\mathrm{SNR}} n(k). \qquad (2.3)$$

Der Faktor β_{SNR} wird so eingestellt, dass sich für den Ausgang $y(k)$ das gewünschte Signal-zu-Geräusch-Verhältnis (SNR) einstellt. Da sowohl Faltung als auch Addition lineare Operationen sind, können mit diesem Ansatz keine nicht-linearen Effekte modelliert werden. In der beschriebene Anordnung treten solche Nichtlinearitäten in Form von Mikrofonverzerrungen auf. Da für die vorliegenden Daten hochwertige Mikrofone benutzt wurden, können die Auswirkungen der fehlenden Möglichkeit der Modellierung von Nichtlinearitäten vernachlässigt werden.

Für die Aufnahmen von Fahrzeuggeräusch und Raumimpulsantworten wurde ein Kraftfahrzeuginnenraum mit mehreren Mikrofonen ausgestattet. Dabei wurden nicht nur der klassische Einbauort am Rückspiegel gewählt, sondern zusätzlich verschiedene andere Positionen wie beispielsweise die A-Säule, die Sonnenblende oder das Schiebedachbedienteil. Die unterschiedlichen Mikrofonpositionen sind in Abbildung 2.10 dargestellt. Mit ihnen können sowohl lineare Mikrofonarrays, darunter versteht man die äquidistante Anordnung von Sensoren auf einer Geraden im Raum oder in der Ebene, als auch nahezu beliebig verteilte Mikrofonanordnungen realisiert werden.

Mit dem so ausgestatteten Fahrzeug wurde für folgende Fahrsituationen das Hintergrundgeräusch aufgezeichnet:

[4] Die Audiodaten wurden von der Firma Harman/Becker in Ulm zur Verfügung gestellt.

[5] Da die hier vorkommenden Signale rein reell sind, wird $h(l)$ in der Faltungssumme nicht konjugiert komplex geschrieben.

- Autobahnfahrt bei verschiedenen konstanten Geschwindigkeiten (100, 130 und 160 km/h),
- Fahrt im Stadtverkehr mit wechselnden Geschwindigkeiten.

Aufgrund ihrer Charakteristik werden die Autobahnaufnahmen für die Analyse der vorgeschlagenen Algorithmen bei stationärem Geräusch benötigt, die Stadtverkehraufnahmen für Tests bei instationärem Geräusch.

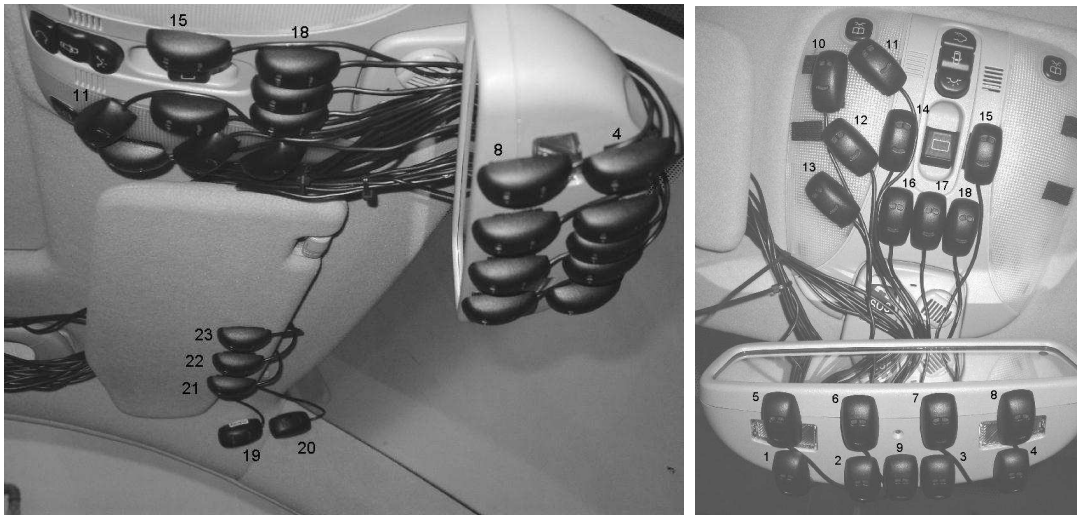

Abbildung 2.10: *Mikrofonanordnung für die Messung der Raumimpulsantworten und des Fahrzeuggeräuschs. [Mit freundlicher Genehmigung der Firma Harman/Becker, Ulm.]*

Die Aufnahmen wurden mit einer maximalen Abtastrate von 16 kHz durchgeführt, wobei diese für die vorliegende Arbeit auf eine Abtastfrequenz von 8 kHz begrenzt wurde. Dies ist eine gebräuchliche Näherung [52] für die bei der Übertragung von Telefonsignalen verwendete *Telefonbandbreite*, bei der ein Bandpass mit den Grenzfrequenzen $f_0 = 300$ Hz und $f_1 = 3,6$ kHz benutzt wird.

Die Messungen der Raumimpulsantworten wurden in einem geschlossenen Raum durchgeführt, um Verfälschungen durch externe Geräuschquellen zu vermeiden. Die verwendete Messanordnung ist in Abbildung 2.11 dargestellt. Auf dem Sitz des Fahrers befindet sich ein Kunstkopf, in dessen Mund ein Lautsprecher mit einer dem menschlichen Mund nachempfundenen Abstrahlcharakteristik eingebaut ist. Mit diesem wird ein künstliches Rauschsignal ausgesendet, welches von den in Abbildung 2.10 dargestellten Mikrofonen, die sich an verschiedenen Positionen innerhalb des Fahrzeuginnenraums befinden, aufgenommen wird. Pro Mikrofon kann aus Referenzsignal und empfangenen Rauschsignal mit Hilfe der Kreuz- und Autoleistungsdichtespektren die gesuchte Raumimpulsantwort im Frequenz-

bereich berechnet werden:

$$H(e^{j\Omega}) \quad = \quad \frac{S_{yy}(e^{j\Omega})}{S_{s_{\mathrm{ref}}y}(e^{j\Omega})}. \tag{2.4}$$

Als Referenzsignal $s_{\mathrm{ref}}(k)$ wird dabei nicht das dem Lautsprecher im Kunstkopf zugeführte Signal verwendet, sondern das mit einem kalibrierten, sich vor dem Kunstkopf im sogenannten *Mund-Referenzpunkt* montierten Mikrofon aufgezeichnete.

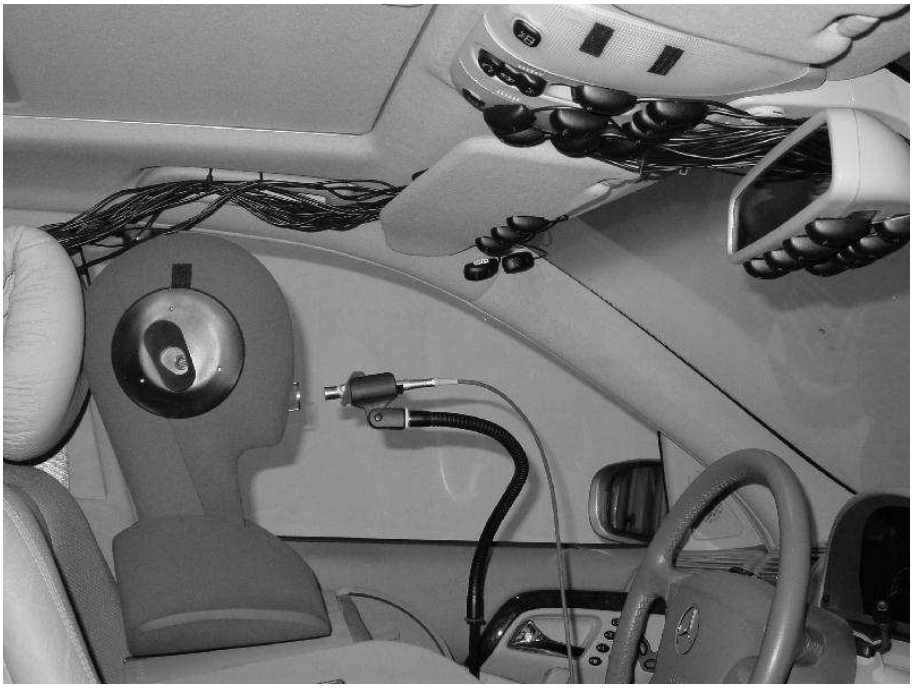

Abbildung 2.11: *Für die Messung der Raumimpulsantworten verwendeter Kunstkopf. [Mit freundlicher Genehmigung der Firma Harman/Becker, Ulm.]*

Für den Kunstkopf wurden dabei verschiedene Einbaupositionen berücksichtigt, um unterschiedliche Eigenschaften des Fahrers zu modellieren:

- **Fahrergröße**: Diese wirkt sich sowohl auf die Sitzposition wie auch auf die Sitzhöhe aus. Beispielsweise bringen kleine Fahrer den Sitz in eine vordere Position, während große Personen eine rückwärtige Einstellung verwenden. Dadurch ergeben sich unterschiedlich lange Ausbreitungswege des direkten Schalls sowie Unterschiede in der Mehrwegeausbreitung.

- **Kopfdrehung**: Wird während des Fahrens telefoniert, so kann nicht davon ausgegangen werden, dass der Fahrer die ganze Zeit in Richtung der Mikrofone spricht. Der Schulterblick beim Spurwechsel macht beispielsweise ein Drehen des Kopfes um bis zu 90 Grad erforderlich. Außerdem sind verteilte Mikrofonanordnungen im Kraftfahrzeug möglich (zum Beispiel A-Säule und Rückspiegel), bei denen es überhaupt nicht möglich ist, gleichzeitig direkt in alle Mikrofone zu sprechen.

Dazu werden drei verschiedene Sitzpositionen (hinten, Mitte, vorne) sowie pro Sitzposition nochmals drei Kopfstellungen (geradeaus, links, rechts) berücksichtigt. Dadurch ergibt sich für jedes Mikrofon ein Satz von neun Raumimpulsantworten, die in Tabelle 2.1 aufgelistet sind. Von den neun dort genannten Szenarien, stellt die hintere Sitzposition (maximale Entfernung zu den Mikrofonen) mit nach links (in die den Mikrofonen abgewandte Richtung) sprechendem Fahrer den schwierigsten Fall (Englisch: Worst Case) dar.

Tabelle 2.1: *Vorhandene Datensätze für die Raumimpulsantwort $h(k)$ jedes Mikrofons.*

Sitzposition	**Kopfausrichtung**
vorne	geradeaus nach links nach rechts
mittig	geradeaus nach links nach rechts
hinten	geradeaus nach links nach rechts

Die Raumimpulsantworten bewirken sowohl eine Verzögerung des ungestörten Sprachsignals aufgrund dessen Ausbreitung von der jeweiligen Sitzposition zum Mikrofon mit Schallgeschwindigkeit ($c_{\text{Schall}} = 330$ m/s) als auch eine Verhallung durch die Mehrwegeausbreitung. Wird beispielsweise mit nach links gedrehtem Kopf gesprochen, so gibt es neben dem direkten Weg zum Mikrofon noch Ausbreitungswege mit Reflexion an der Türscheibe der Fahrerseite. Eventuell vorhandene, durch die Übertragungskennlinien realer Mikrofone verursachte lineare Verzerrungen werden ebenfalls durch die Raumimpulsantwort abgebildet.

Die Auswirkungen der Verhallung des ungestörten Sprachsignals aus Abbildung 2.2 durch die Raumimpulsantwort im Worst Case (nach links sprechender, großer Fahrers) ist in Abbildung 2.12 dargestellt. Das durch die Überlagerung der einzelnen Reflexionen verursachte Verschmieren der Feinstruktur ist besonders beim letzten Sprachblock deutlich zu erkennen.

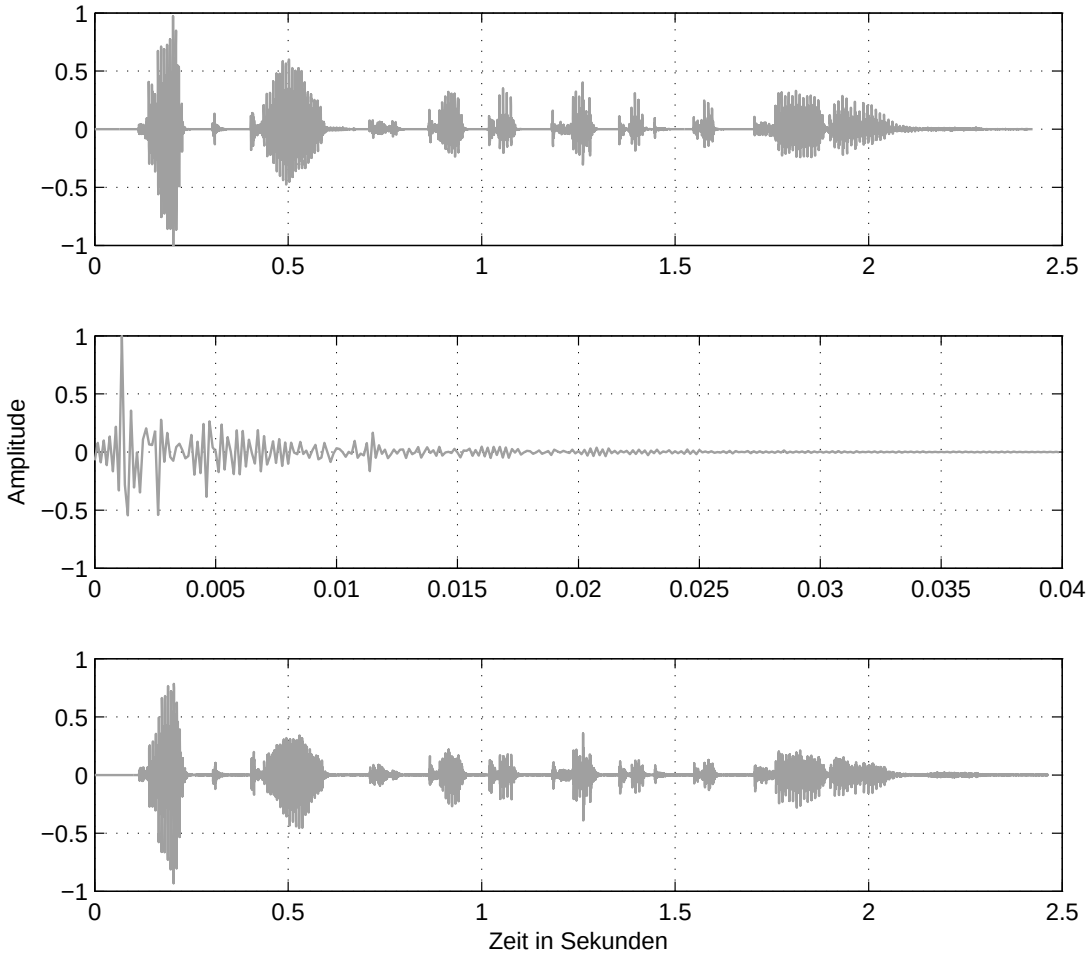

Abbildung 2.12: *Das ungestörte Sprachsignal (oben) wird durch die Faltung mit der Raumimpulsantwort (Mitte) verhallt. Das resultierende Signal (unten) weist eine veränderte Feinstruktur auf.*

Mit den zuvor beschriebenen Daten ist es nun möglich, Sprachsignale innerhalb eines Kraftfahrzeuges zu simulieren, indem die einzelnen Mikrofonsignale $y_i(k)$ mit $i=1,\ldots,N$ mittels Gleichung 2.3 berechnet werden. Darüber hinaus können auch Sonderfälle wie zum Beispiel die identische Verhallung aller Mikrofone oder die Drehung des Fahrerkopfes (durch sprunghaftes Umschalten der Raumimpulsantwort) generiert werden. Als Ausgangsdaten wurden dabei Signale unterschiedlicher Sprachdatenbanken wie zum Beispiel der *TIMIT – Acoustic-Phonetic Continuous Speech Corpus* verwendet.

Kapitel 3

Signalmodell und Kalman-Filter

Kern der vorgestellten Verfahren zur akustischen Geräuschreduktion ist ein speziell auf die Verwendung zur Geräuschreduktion in einer Freisprecheinrichtung angepasstes, mehrkanaliges Kalman-Filter[1], welches im Folgenden hergeleitet wird. Hierzu ist dieses Kapitel in fünf Abschnitte geteilt.

Zuerst wird das Signalmodell, welches für die Ableitung des Kalman-Filters notwendig ist, motiviert und im Zustandsraum beschrieben. Der zweite Teil beschäftigt sich mit der eigentlichen Herleitung der mehrkanaligen Kalman-Filtergleichungen. Zwecks besserer Lesbarkeit der Gleichungen geschieht dies zunächst für den Fall von zwei Mikrofonen (Kanälen), bevor im darauf folgenden Abschnitt die Erweiterung auf mehr als zwei Kanäle diskutiert wird. Der vierte Teil befasst sich mit der Interpretation der Kalman-Filtergleichungen. Dabei wird insbesondere die Messgleichung analysiert und deren Wirkung im Gesamtalgorithmus veranschaulicht. Im letzten Abschnitt werden Kalman-Filter verschiedener Kanalanzahl und verschiedener Ordnung anhand des numerischen Aufwands miteinander verglichen.

Obwohl das spätere Verfahren im Teilbandbereich implementiert wird, wird die folgende Herleitung im Vollband durchgeführt, um die Notation der einzelnen Größen nicht durch Angabe des Frequenzbandindex' μ unnötig zu verkomplizieren. Allerdings werden im Hinblick auf besagte Teilbandimplementierung explizit komplexwertige Eingangssignale angenommen, so dass die hier gefundenen Gleichungen später direkt für jedes Teilband ausgeführt werden können.

Die Überlegungen dieses Kapitels beschränken sich auf den eigentlichen Kalman-Filter-Algorithmus. Für dessen Verwendung ist die fortlaufende Schätzung einiger Parameter notwendig. Mit diesen und den zugehörigen Schätzverfahren befasst sich Kapitel 4.

[1] Mit dem Begriff *Kalman-Filter* ist hier grundsätzlich der nach R. E. Kalman benannte Algorithmus [25] eines stochastischen Zustandschätzers gemeint.

3.1 Signalmodell

Für die Herleitung des Kalman-Filters ist stets ein Signalmodell erforderlich [17]. Im Kontext der vorliegenden Arbeit muss dieses die akustischen Vorgänge im Fahrzeuginnenraum nachbilden. Auf Basis dieses Modells kann danach ein Kalman-Filter hergeleitet werden, dessen Struktur eine Kopie des zugrunde liegenden Modells darstellt.

In diesem Abschnitt wird zunächst die Modellierung, welche für den Fahrzeuginnenraum und die darin vorkommenden Signale gewählt wurde, diskutiert. Danach wird die mathematische Beschreibung des Modells im Zeitbereich und im Zustandsraum vorgestellt. Aus dieser werden im darauf folgenden Abschnitt die Gleichungen des Kalman-Filters hergeleitet (siehe auch [27]).

3.1.1 Quellensignal- und Fahrzeuginnenraummodellierung

Im Hinblick auf die spätere Beschreibung im Zustandsraum mittels System- und Messgleichung erweist sich eine Aufteilung der Modellierung in *Quellensignal-* und *Fahrzeuginnenraummodellierung* als sinnvoll. Erstere beinhaltet die Entstehung der Sprach- und Geräuschkomponenten, letztere deren Zusammenfügen unter Berücksichtigung des Einflusses des Fahrzeuginnenraums.

Beschreibung der Quellensignale

Die im Fahrzeuginnenraum auftretenden akustischen Signale - Sprache und Fahrzeuggeräusch - wurden bereits in Kapitel 2 hinsichtlich ihrer stochastischen Eigenschaften beschrieben.

Aufgrund der zeitweise harmonischen Struktur von Sprache bietet sich zur deren Beschreibung das autoregressives Modell (AR-Modell) an. Dessen Parameter müssen allerdings regelmäßig neu geschätzt werden, da Stationarität bei Sprachsignalen nur für Intervalle bis zu einer Dauer von ca. 20 bis 50 ms angenommen werden kann (siehe Abschnitt 2.1.1).

Dagegen weist Fahrzeuggeräusch im Allgemeinen eine glattere, nicht durch eine harmonische Struktur geprägte, spektrale Einhüllende auf. Sieht man zusätzlich von transienten Ereignissen ab, wie sie z.B. beim Wechsel des Fahrbahnbelags auftreten können, so ändern sich die statistischen Eigenschaften verglichen mit denen von Sprache nur sehr langsam. Aus Gründen der Einfachheit wird daher ebenfalls das AR-Modell gewählt. Die Wahl der Ordnungen für Sprache und Fahrzeuggeräusch wird in Abschnitt 5.2.1 diskutiert.

Beschreibung des Fahrzeuginnenraums

Für die anschließende Modellierung des Fahrzeuginnenraums werden folgende Annahmen getroffen:

- Verwendung eines Arrays mit zunächst zwei Mikrofonen[2].
- Es spricht immer nur die auf dem Fahrersitz sitzende Person, d.h. es gibt nur eine Sprachquelle im Fahrzeug.
- An den Mikrofonen aufgenommenen Hintergrundgeräusche weisen sowohl zueinander korrelierte als auch unkorrelierte Anteile auf.

Zusammengefasst bedeutet dies, dass drei Quellensignale von zwei Mikrofonen aufgenommen werden. Das jeweilige Mikrofonsignal setzt sich also durch Superposition folgender drei Komponenten zusammen:

1. Durch den Raum verhalltes Sprachsignal.
2. Zu anderem Mikrofonsignal unkorreliertes Geräusch.
3. Zu anderem Mikrofonsignal korreliertes Geräusch.

Diese zunächst grobe Modellbeschreibung ist im Signalflussdiagramm aus Abbildung 3.1 bereits ersichtlich. Eine detaillierte Herleitung und Betrachtung der einzelnen Blöcke wird im weiteren Verlauf dieses Kapitels vorgestellt.

3.1.2 Modellierung im Zeitbereich

Im Folgenden werden für die Sprach- bzw. Geräuschmodellierung AR-Modelle der Ordnung p bzw. q definiert. Diese werden so formuliert, dass deren Koeffizienten denen des linearen Prädiktors entsprechen, weshalb in diesem Kapitel zur Kenntlichmachung stets von Prädiktorkoeffizienten gesprochen wird. Der Zusammenhang zwischen AR-Modellierung und linearer Prädiktion wird in Abschnitt 4.1.3 beschrieben, die Wahl der Modellordnungen in Abschnitt 5.2.1. Unter der bereits getroffenen Annahme, dass niemals zwei Personen im KFZ gleichzeitig sprechen, ergibt sich für das Signal der Sprachquelle $s(k)$ am Mund des Sprechers:

$$s(k) \;=\; \sum_{i=1}^{p} a_{s,i}(k-1)s(k-i) + v(k). \tag{3.1}$$

Hierbei beschreibt $a_{s,i}(k)$ den i-ten zeitvarianten Prädiktorkoeffizient der Sprachkomponente zum Abtastzeitpunkt k. Das Anregungssignal $v(k)$ des Sprachmo-

[2]Dies schränkt die Allgemeinheit nicht ein. Eine Erweiterung auf mehr als zwei Mikrofone wird in 3.3 behandelt.

dells ist statistisch unabhängiges, weißes Rauschen. Im Gegensatz zur Faltung (siehe beispielsweise Gleichungen 3.5 und 3.6) werden die AR-Modelle nicht konjugiert-komplex angesetzt.

Da die an den Mikrofonen aufgenommenen Geräusche sowohl zueinander korrelierte als auch unkorrelierte Anteile aufweisen, werden für die Modellierung bei zwei Mikrofonen mindestens zwei Geräuschquellen benötigt. Analog zu Gleichung 3.1 ergeben sich für die beiden Quellensignale $n_1(k)$ und $n_2(k)$:

$$n_1(k) = \sum_{i=1}^{q} a_{n_1,i}(k-1)n_1(k-i) + w_1(k), \tag{3.2}$$

$$n_2(k) = \sum_{i=1}^{q} a_{n_2,i}(k-1)n_2(k-i) + w_2(k), \tag{3.3}$$

wobei $w_1(k)$ und $w_2(k)$ wiederum als statistisch unabhängige, weiße Rauschprozesse definiert werden. Daher gilt für die zugehörigen Kreuzkorrelationsfunktionen:

$$r_{vw_1}(k,l) = r_{vw_2}(k,l) = r_{w_1w_2}(k,l) = 0 \quad \text{für} \quad \forall\, k,l. \tag{3.4}$$

Die Mikrofonsignale werden nun durch Filterung der drei Quellensignale $s(k)$, $n_1(k)$ und $n_2(k)$ mit anschließender additiver Überlagerung generiert.

Zunächst wird die Sprachkomponente am jeweiligen Mikrofon als Faltung der Quelle $s(k)$ mit der zeitvarianten Raumimpulsantwort $h_{1,i}(k)$ bzw. $h_{2,i}(k)$ modelliert. Für die unkorrelierten Geräuschkomponenten werden die Quellensignale $n_1(k)$ und $n_2(k)$ direkt verwendet, während die zueinander korrelierten Anteile durch Faltung mit den ebenfalls zeitvarianten Filterfunktionen $g_{21,i}(k)$ bzw. $g_{12,i}(k)$ beschrieben werden. Dadurch ergibt sich für die Mikrofonsignale $y_1(k)$ und $y_2(k)$:

$$y_1(k) = \sum_{i=0}^{p-1} h^*_{1,i}(k)s(k-i) + n_1(k) + \sum_{i=0}^{q-1} g^*_{21,i}(k)n_2(k-i), \tag{3.5}$$

$$y_2(k) = \sum_{i=0}^{p-1} h^*_{2,i}(k)s(k-i) + n_2(k) + \sum_{i=0}^{q-1} g^*_{12,i}(k)n_1(k-i). \tag{3.6}$$

Die vordere Summe repräsentiert dabei die Sprachkomponente, die hintere Summe die korrelierte und der mittlere Term die unkorrelierte Geräuschkomponente innerhalb des jeweiligen Mikrofonsignals.

Das gesamte Modell mit Anregungs-, Quellen- und Mikrofonsignalen ist in Abbildung 3.1 als Signalflussdiagramm dargestellt. Die Superposition der drei Signalkomponenten (Sprache, unkorreliertes und korreliertes Geräusch) ist darin deutlich erkennbar. Aus Gründen der Übersicht wurden die Impulsantworten als zeitinvariant abgebildet.

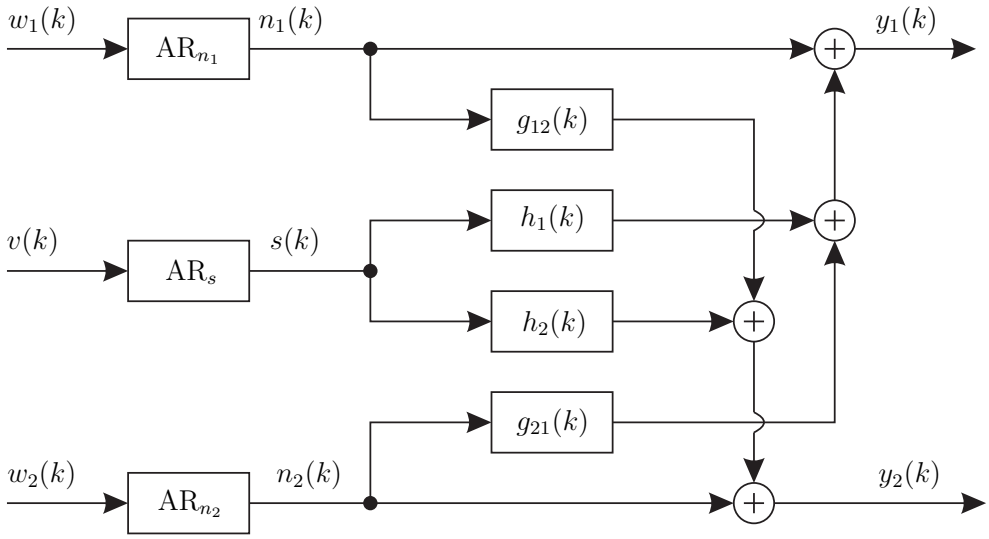

Abbildung 3.1: *Signalmodell im Zeitbereich.*

3.1.3 Modellierung im Zustandsraum

Um zu der Zustandsraumdarstellung zu gelangen, die für die Herleitung des Kalman-Filters notwendig ist, werden die bis jetzt gefunden Gleichungen in Vektor/Matrix-Schreibweise übersetzt. Dazu werden folgende Spaltenvektoren definiert:

$$\mathbf{s}(k) = [s(k-p+1), s(k-p+2), \ldots, s(k-1), s(k)]^{\mathrm{T}}, \tag{3.7}$$

$$\mathbf{n}_1(k) = [n_1(k-q+1), n_1(k-q+2), \ldots, n_1(k-1), n_1(k)]^{\mathrm{T}}, \tag{3.8}$$

$$\mathbf{n}_2(k) = [n_2(k-q+1), n_2(k-q+2), \ldots, n_2(k-1), n_2(k)]^{\mathrm{T}}. \tag{3.9}$$

Man beachte, dass der am weitesten in der Vergangenheit liegende Abtastwert jeweils oben, der aktuelle Abtastwert jeweils unten im Vektor steht. Dieses Festlegung ist nicht zwingend und orientiert sich an [18, 19]. Die umgekehrte Darstellung mit dem aktuellsten Abtastwert oben findet sich beispielsweise in [21]. Fasst man die als Zeilenvektor geschriebenen Sprach-Prädiktorkoeffizienten mit einer Schiebematrix zusammen, so entsteht die sogenannte *Transitions-* oder **A**-Matrix des Sprachmodells:

$$\mathbf{A}_s(k|k-1) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ a_{s,p}(k-1) & a_{s,p\text{-}1}(k-1) & a_{s,p\text{-}2}(k-1) & \cdots & a_{s,1}(k-1) \end{bmatrix}. \tag{3.10}$$

Die Notation des Zeitindex' $k|k-1$ sagt aus, dass ein Übergang (eine Transition) des Sprachsignalvektors $\mathbf{s}$ vom Abtastzeitpunkt $k-1$ zum Abtastzeitpunkt k

durchgeführt wird.

Analog wird mit den Prädiktorkoeffizienten der beiden Geräuschquellen verfahren. Die Quellensignale (Gleichungen 3.1-3.3) lassen sich somit folgendermaßen schreiben:

$$\mathbf{s}(k) = \mathbf{A}_s(k|k-1)\mathbf{s}(k-1) + [\ |\ ]v(k), \tag{3.11}$$
$$\mathbf{n}_1(k) = \mathbf{A}_{n1}(k|k-1)\mathbf{n}_1(k-1) + [\ |\ ]w_1(k), \tag{3.12}$$
$$\mathbf{n}_2(k) = \mathbf{A}_{n2}(k|k-1)\mathbf{n}_2(k-1) + [\ |\ ]w_2(k). \tag{3.13}$$

Dabei stellt $[\ |\ ]$ die symbolische Abkürzung für einen Spaltenvektor der Form $[0, 0, \ldots, 0, 1]^{\mathrm{T}}$ und entsprechender Länge (*Ausschneidevektor*) dar. Die *Schiebematrix* besteht aus den ersten $p-1$ bzw. $q-1$ Zeilen der jeweiligen $\mathbf{A}$-Matrix und sorgt dafür, dass beim Übergang vom Abtastzeitpunkt $k-1$ zum Abtastzeitpunkt k der älteste Abtastwert $s(k-p+1)$ nach oben aus dem Vektor $\mathbf{s}(k)$ hinausgeschoben wird. Die am unteren Ende des Vektors entstandene Lücke wird mit dem Wert $s(k)$, der aus den Prädiktorkoeffizienten und dem Anregungssignal $v(k)$ berechnet wird, aufgefüllt.

Das gleiche Vorgehen wird nun auf die Mikrofonsignale $y_1(k)$ und $y_2(k)$ angewendet. Wiederum werden die Faltungssummen durch Skalarprodukte ersetzt. Dazu werden folgende Spaltenvektoren definiert ($\mathbf{h}_2(k)$ und $\mathbf{g}_{12}(k)$ entsprechend):

$$\mathbf{h}_1(k) = [h_{1,p-1}(k), h_{1,p-2}(k), \ldots, h_{1,1}(k), h_{1,0}(k)]^{\mathrm{T}}, \tag{3.14}$$
$$\mathbf{g}_{21}(k) = [g_{21,q-1}(k), g_{21,q-2}(k), \ldots, g_{21,1}(k), g_{21,0}(k)]^{\mathrm{T}}. \tag{3.15}$$

Gleichungen 3.5 und 3.6 lassen sich damit in Vektor/Matrix-Schreibweise darstellen:

$$y_1(k) = \mathbf{h}_1^{\mathrm{H}}(k)\mathbf{s}(k) + [\ |\ ]^{\mathrm{T}}\,\mathbf{n}_1(k) + \mathbf{g}_{21}^{\mathrm{H}}(k)\mathbf{n}_2(k), \tag{3.16}$$
$$y_2(k) = \mathbf{h}_2^{\mathrm{H}}(k)\mathbf{s}(k) + \mathbf{g}_{12}^{\mathrm{H}}(k)\mathbf{n}_1(k) + [\ |\ ]^{\mathrm{T}}\,\mathbf{n}_2(k). \tag{3.17}$$

Die Geräuschvektoren $\mathbf{n}_1(k)$ und $\mathbf{n}_2(k)$ in Gleichung 3.17 wurden so umgestellt, dass sie in beiden Gleichungen jeweils untereinander stehen.

Die vollständige Zustandsraumdarstellung ist in Abbildung 3.2 dargestellt. Aus Gründen der Übersichtlichkeit sind alle Größen ohne den Zeitindex k notiert. Fettgedruckte Pfeile bezeichnen vektorielle, dünngedruckte skalare Größen.

Um zu den Ausgangsgleichungen für die Herleitung des Kalman-Filters zu gelangen, werden die Gleichungen 3.11 bis 3.13 sowie 3.16 und 3.17 weiter vereinfacht. Zunächst werden die drei Signalvektoren $\mathbf{s}(k)$,$\mathbf{n}_1(k)$ und $\mathbf{n}_2(k)$ in dem *Zustandsvektor* $\mathbf{x}(k)$ sowie die drei Anregungssignale $v(k)$, $w_1(k)$ und $w_2(k)$ in dem *Anregungsvektor* $\mathbf{u}(k)$ zusammengefasst:

$$\mathbf{x}(k) = \begin{bmatrix} \mathbf{s}(k) \\ \mathbf{n}_1(k) \\ \mathbf{n}_2(k) \end{bmatrix}, \qquad \mathbf{u}(k) = \begin{bmatrix} v(k) \\ w_1(k) \\ w_2(k) \end{bmatrix}. \tag{3.18}$$

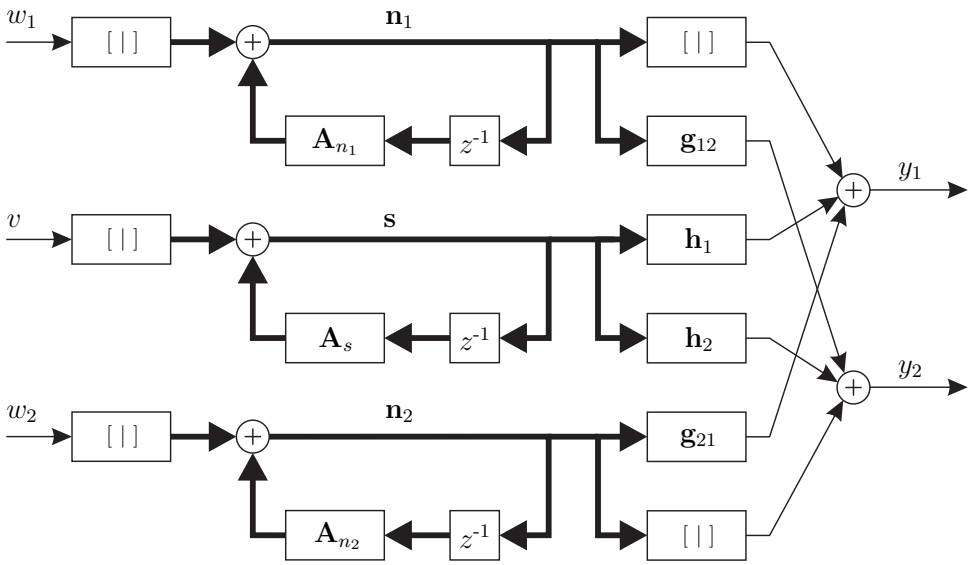

Abbildung 3.2: *Signalmodell in Zustandsraumdarstellung.*

Danach wird eine zusammengesetzte Transitionsmatrix $\mathbf{A}(k|k-1)$ definiert, auf deren Hauptdiagonalen blockweise die drei $\mathbf{A}$-Matrizen $\mathbf{A}_s(k|k-1)$, $\mathbf{A}_{n_1}(k|k-1)$ und $\mathbf{A}_{n_2}(k|k-1)$ angeordnet sind:

$$\mathbf{A}(k|k-1) \;=\; \begin{bmatrix} \mathbf{A}_s(k|k-1) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{n_1}(k|k-1) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{n_2}(k|k-1) \end{bmatrix}. \tag{3.19}$$

Die fettgedruckte Null $\mathbf{0}$ bezeichnet hierbei eine Matrix bzw. einen Vektor entsprechender Größe mit allen Elementen identisch Null.

Fasst man schließlich die Ausschneidevektoren $[\,|\,]$ in gleicher Weise in der Matrix $\mathbf{B}$ zusammen, ergibt sich für die Systemgleichung:

$$\underbrace{\begin{bmatrix} \mathbf{s}(k) \\ \mathbf{n}_1(k) \\ \mathbf{n}_2(k) \end{bmatrix}}_{\mathbf{x}(k)} = \underbrace{\begin{bmatrix} \mathbf{A}_s(k|k-1) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{n_1}(k|k-1) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{n_2}(k|k-1) \end{bmatrix}}_{\mathbf{A}(k|k-1)} \cdot \underbrace{\begin{bmatrix} \mathbf{s}(k\text{-}1) \\ \mathbf{n}_1(k\text{-}1) \\ \mathbf{n}_2(k\text{-}1) \end{bmatrix}}_{\mathbf{x}(k-1)} \cdots$$
$$+ \underbrace{\begin{bmatrix} [\,|\,] & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & [\,|\,] & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & [\,|\,] \end{bmatrix}}_{\mathbf{B}} \cdot \underbrace{\begin{bmatrix} v(k) \\ w_1(k) \\ w_2(k) \end{bmatrix}}_{\mathbf{u}(k)}. \tag{3.20}$$

Durch ähnliches Vorgehen kann man die Messgleichung bestimmen. Hierzu werden die Mikrofonsignale $y_1(k)$ und $y_2(k)$ im *Messvektor* $\mathbf{y}(k)$, die Raumimpulsantworten $\mathbf{h}_1(k)$ und $\mathbf{h}_2(k)$ sowie die Filtervektoren $\mathbf{g}_{12}(k)$ und $\mathbf{g}_{21}(k)$ in der

Messmatrix $\mathbf{C}(k)$ zusammengefasst:

$$\mathbf{y}(k) = \begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix}, \qquad \mathbf{C}(k) = \begin{bmatrix} \mathbf{h}_1(k) & \mathbf{h}_2(k) \\ [\,|\,] & \mathbf{g}_{12}(k) \\ \mathbf{g}_{21}(k) & [\,|\,] \end{bmatrix}. \tag{3.21}$$

Danach kann bereits die Messgleichung notiert werden:

$$\underbrace{\begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix}}_{\mathbf{y}(k)} = \underbrace{\begin{bmatrix} \mathbf{h}_1^{\mathrm{H}}(k) & [\,|\,]^{\mathrm{T}} & \mathbf{g}_{21}^{\mathrm{H}}(k) \\ \mathbf{h}_2^{\mathrm{H}}(k) & \mathbf{g}_{12}^{\mathrm{H}}(k) & [\,|\,]^{\mathrm{T}} \end{bmatrix}}_{\mathbf{C}^{\mathrm{H}}(k)} \cdot \underbrace{\begin{bmatrix} \mathbf{s}(k) \\ \mathbf{n}_1(k) \\ \mathbf{n}_2(k) \end{bmatrix}}_{\mathbf{x}(k)}. \tag{3.22}$$

Zusammenfassend erhält man die Systemgleichung und die Messgleichung mit jeweils zeitvarianten Parametern:

$$\mathbf{x}(k) = \mathbf{A}(k|k-1)\mathbf{x}(k-1) + \mathbf{B}\mathbf{u}(k) \tag{3.23}$$

$$\mathbf{y}(k) = \mathbf{C}^{\mathrm{H}}(k)\mathbf{x}(k). \tag{3.24}$$

Diese dienen als Ausgangsbasis für die Herleitung der eigentlichen Kalman-Filtergleichungen.

3.2 Multiple-Input Single-Output (MISO) Kalman-Filter

Das Kalman-Filter ist eine rekursive Rechenvorschrift, die eine Schätzung des nur gestört messbaren Zustands $\mathbf{x}(k)$ eines linearen, zeitvarianten Systems berechnet. Der so erhaltene Schätzwert ist linear, erwartungstreu und weist eine minimale Fehlervarianz (Optimalfilter) auf [19, 21].

Als Vergleich dazu kann das Wiener-Filter angeführt werden, welches ebenfalls ein Nutzsignal optimal im Sinne des mittleren quadratischen Fehlers von einer additiven Störung trennt. Allerdings gilt dessen Optimalität nur unter der Annahme stationärer Eingangssignale und nur, wenn sich das Filter im eingeschwungenen Zustand befindet. Bei der Verwendung instationärer Signale (z.B. Sprache) verliert es seine Optimalität. Für diesen Fall existieren Verallgemeinerungen des Wiener-Filters, die auf der Verwendung von Kurzzeitspektren basieren [18, 52, 53]. Durch Anwendung des Kalman-Filters umgeht man hingegen dieses Problem vollständig, da dessen Ausgang ohne Einschwingen sofort optimal ist und keine Stationarität angenommen werden muss.

3.2.1 Ansatz

Der Ansatz geht von den im letzten Abschnitt formulierten Gleichungen 3.23 und 3.24 aus. Um die Herleitung allgemein zu halten, wird die Messgleichung um den Pseudo-Rauschvektor $\mathbf{z}(k) = [z_1(k),\ z_2(k)]^{\mathrm{T}}$ (Messrauschen) ergänzt, der später zur Steuerung der Robustheit verwendet wird:

$$\mathbf{y}(k) \quad = \quad \mathbf{C}^{\mathrm{H}}(k)\mathbf{x}(k) + \mathbf{z}(k). \tag{3.25}$$

Die beiden Rauschprozesse $z_1(k)$ und $z_2(k)$ sind statistisch unabhängig und weiß. Darüber hinaus werden noch folgende Annahmen getroffen, die keine Auswirkungen auf die Praxis haben [18, 21, 17]:

- Das Verfahren beginnt zum Abtastzeitpunkt $k = 0$.
- Der Anfangswert des Zustandsvektors ist mittelwertfrei $\mathrm{E}\{\mathbf{x}(0)\} = \mathbf{0}$ und dessen Kovarianzmatrix $\mathbf{P}_{\mathbf{x}}(0) = \mathrm{E}\{\mathbf{x}(0)\mathbf{x}^{\mathrm{H}}(0)\}$ bekannt.
- Das Anregungssignal $\mathbf{u}(k)$ ist mittelwertfrei $\mathrm{E}\{\mathbf{u}(k)\} = \mathbf{0}$ und weiß, d.h. die zugehörige Kovarianzmatrix $\mathbf{P}_{\mathbf{u}}(k) = \mathrm{E}\{\mathbf{u}(k)\mathbf{u}^{\mathrm{H}}(k)\}$ enthält nur auf der Hauptdiagonalen Elemente ungleich Null.
- Der Anfangszustand $\mathbf{x}(0)$, der Anregungsprozess $\mathbf{u}(k)$ sowie das Messrauschen $\mathbf{z}(k)$ sind jeweils zueinander unkorreliert für alle k:

  1.) $\mathrm{E}\{\mathbf{x}(0)\mathbf{u}^{\mathrm{H}}(k)\} = \mathbf{0}$,

  2.) $\mathrm{E}\{\mathbf{x}(0)\mathbf{z}^{\mathrm{H}}(k)\} = \mathbf{0}$,

  3.) $\mathrm{E}\{\mathbf{u}(k)\mathbf{z}^{\mathrm{H}}(k)\} = \mathbf{0}$.
- Die Transitionsmatrix $\mathbf{A}(k|k-1)$ ist für $k > 0$, die Messmatrix $\mathbf{C}(k)$ für $k \geq 0$ bekannt. Beiden werden daher als determinierte Größen betrachtet.

Die Gleichungen des Kalman-Filters werden in drei Schritten hergeleitet: Dabei beschreiben die ersten beiden die eigentliche Rekursion, welche sich in *Prädiktion* und *Korrektur* aufteilt. Der letzte Schritt beinhaltet die *Initialisierung* des Verfahrens.

3.2.2 Prädiktion

Zunächst wird angenommen, dass ein erwartungstreuer und optimaler Schätzwert $\hat{\mathbf{x}}(k-1|k-1)$ für den Zustandsvektor zum Abtastzeitpunkt $k-1$ (erster Zeitindex) auf Grundlage aller Eingangsdaten bis zum Abtastzeitpunkt $k-1$ (zweiter Zeitindex) existiert.

Ausgehend davon wird nun ein a-priori Schätzwert $\hat{\mathbf{x}}(k|k-1)$ für den Abtastzeit-

punkt k aus den bisherigen Eingangsdaten $\mathbf{y}(l)$, $0 \leq l \leq k-1$ und der letzten Schätzung $\hat{\mathbf{x}}(k{-}1|k{-}1)$ prädiziert. Als linearer Ansatz wird die homogene Lösung der Systemgleichung gewählt:

$$\hat{\mathbf{x}}(k|k-1) = \mathbf{A}(k|k-1)\hat{\mathbf{x}}(k-1|k-1). \tag{3.26}$$

Das Anregungssignal $\mathbf{u}(k)$ ist weiß und geht somit nicht in die Prädiktion ein.

Aus der Mittelwertfreiheit von Anfangszustand $\mathbf{x}(0)$ und Anregungssignal $\mathbf{u}(k)$ ergibt sich, dass alle nachfolgenden Zustände $\mathbf{x}(k)$ für $k > 0$ ebenfalls mittelwertfrei sind. Für Erwartungstreue muss also $\mathrm{E}\{\hat{\mathbf{x}}(k|k{-}1)\} = \mathbf{0}$ gelten. Da $\hat{\mathbf{x}}(k{-}1|k{-}1)$ wie oben erwähnt ein erwartungstreuer Schätzwert ist, gilt: $\mathrm{E}\{\hat{\mathbf{x}}(k{-}1|k{-}1)\} = \mathbf{0}$. Durch Anwenden des Erwartungswertoperators auf Gleichung 3.26 erhält man:

$$\mathrm{E}\left\{\hat{\mathbf{x}}(k|k-1)\right\} = \mathbf{0}. \tag{3.27}$$

Der a-priori Schätzwert $\hat{\mathbf{x}}(k|k-1)$ ist somit erwartungstreu.

Für die weitere Herleitung werden zwei Fehlervektoren und ihre Kovarianzmatrizen definiert. Der erste bezeichnet den a-priori, der zweite den a-posteriori Schätzfehler:

$$\begin{aligned}
\mathbf{e}(k|k-1) &= \mathbf{x}(k) - \hat{\mathbf{x}}(k|k-1), && (3.28)\\
\mathbf{e}(k|k) &= \mathbf{x}(k) - \hat{\mathbf{x}}(k|k), && (3.29)\\
\mathbf{P_e}(k|k-1) &= \mathrm{E}\{\mathbf{e}(k|k-1)\mathbf{e}^{\mathrm{H}}(k|k-1)\}, && (3.30)\\
\mathbf{P_e}(k|k) &= \mathrm{E}\{\mathbf{e}(k|k)\mathbf{e}^{\mathrm{H}}(k|k)\}. && (3.31)
\end{aligned}$$

Setzt man in Gleichung 3.28 die Gleichungen 3.23 und 3.26 ein, erhält man für den a-priori Schätzfehler $\mathbf{e}(k|k-1)$:

$$\begin{aligned}
\mathbf{e}(k|k-1) &= \mathbf{x}(k) - \hat{\mathbf{x}}(k|k-1)\\
&= \mathbf{A}(k|k-1)\left(\mathbf{x}(k-1) - \hat{\mathbf{x}}(k-1|k-1)\right) + \mathbf{B}\mathbf{u}(k)\\
&= \mathbf{A}(k|k-1)\mathbf{e}(k-1|k-1) + \mathbf{B}\mathbf{u}(k). && (3.32)
\end{aligned}$$

Durch Einsetzen in Gleichung 3.30 lässt sich dessen Kovarianzmatrix $\mathbf{Pe}(k|k{-}1)$ bestimmen:

$$\begin{aligned}
\mathbf{P_e}(k|k-1) &= \mathrm{E}\Big\{\left[\mathbf{A}(k|k-1)\mathbf{e}(k-1|k-1) + \mathbf{B}\mathbf{u}(k)\right]\cdots\\
&\quad \left[\mathbf{A}(k|k-1)\mathbf{e}(k-1|k-1) + \mathbf{B}\mathbf{u}(k)\right]^{\mathrm{H}}\Big\}\\
&= \mathbf{A}(k|k-1)\mathrm{E}\left\{\mathbf{e}(k-1|k-1)\mathbf{e}^{\mathrm{H}}(k-1|k-1)\right\}\mathbf{A}^{\mathrm{H}}(k|k-1)\cdots\\
&\quad +\mathbf{B}\mathrm{E}\left\{\mathbf{u}(k)\mathbf{u}^{\mathrm{H}}(k)\right\}\mathbf{B}^{\mathrm{T}}\cdots\\
&\quad +\mathbf{A}(k|k-1)\underbrace{\mathrm{E}\left\{\mathbf{e}(k-1|k-1)\mathbf{u}^{\mathrm{H}}(k)\right\}}_{=\mathbf{0}}\mathbf{B}^{\mathrm{T}}\cdots\\
&\quad +\mathbf{B}\underbrace{\mathrm{E}\left\{\mathbf{u}(k)\mathbf{e}^{\mathrm{H}}(k-1|k-1)\right\}}_{=\mathbf{0}}\mathbf{A}^{\mathrm{H}}(k|k-1).
\end{aligned}$$

Die Erwartungswerte in den zwei letzten Zeilen sind identisch null, da $\mathbf{e}(k{-}1|k{-}1)$ und $\mathbf{u}(k)$ mittelwertfrei und zueinander unkorreliert sind. Dies wird am Beispiel des ersten der zwei Terme gezeigt. Zunächst wird der Schätzfehler $\mathbf{e}(k-1|k-1)$ wieder als Differenz geschrieben:

$$\mathrm{E}\left\{\mathbf{e}(k-1|k-1)\mathbf{u}^{\mathrm{H}}(k)\right\} = \underbrace{\mathrm{E}\left\{\mathbf{x}(k-1)\mathbf{u}^{\mathrm{H}}(k)\right\}}_{(1)} - \underbrace{\mathrm{E}\left\{\hat{\mathbf{x}}(k-1|k-1)\mathbf{u}^{\mathrm{H}}(k)\right\}}_{(2)}.$$

Setzt man in Ausdruck (1) für $\mathbf{x}(k-1)$ die Systemgleichung 3.23 ein, so erhält man:

$$\mathbf{A}(k-1|k-2)\mathrm{E}\left\{\mathbf{x}(k-2)\mathbf{u}^{\mathrm{H}}(k)\right\} + \mathbf{B}\underbrace{\mathrm{E}\left\{\mathbf{u}(k-1)\mathbf{u}^{\mathrm{H}}(k)\right\}}_{=\mathbf{0}}.$$

Der Term $\mathrm{E}\{\mathbf{u}(k-1)\mathbf{u}^{\mathrm{H}}(k)\} = \mathbf{R}_{\mathbf{uu}}(k-1,k)$ verschwindet, da $\mathbf{u}(k)$ weiß ist. Wiederholt man das Vorgehen weitere $(k-2)$-mal, ergibt sich für (1) schließlich:

$$\mathbf{A}(1|0)\mathrm{E}\left\{\mathbf{x}(0)\mathbf{u}^{\mathrm{H}}(k)\right\} + \mathbf{B}\mathrm{E}\left\{\mathbf{u}(1)\mathbf{u}^{\mathrm{H}}(k)\right\} = 0,$$

da Unkorreliertheit zwischen dem Anfangszustand $\mathbf{x}(0)$ und dem Anregungssignal $\mathbf{u}(k)$ angenommen wurde.

Ausdruck (2) verschwindet, da $\mathbf{u}(k)$ aufgrund Gleichung 3.23 nur auf $\mathbf{x}(k)$ wirkt, nicht aber auf $\mathbf{x}(k{-}1)$, $\mathbf{x}(k{-}2)$, usw. Das Anregungssignal $\mathbf{u}(k)$ ist daher unkorreliert zu dem Schätzwert einer dieser Zustände (hier: $\hat{\mathbf{x}}(k{-}1|k{-}1)$).

Nach diesen Überlegungen ergibt sich die Kovarianzmatrix des a-priori Schätzfehlers zu:

$$\mathbf{P}_{\mathbf{e}}(k|k-1) = \mathbf{A}(k|k-1)\mathbf{P}_{\mathbf{e}}(k-1|k-1)\mathbf{A}^{\mathrm{H}}(k|k-1) + \mathbf{B}\mathbf{P}_{\mathbf{u}}(k)\mathbf{B}^{\mathrm{T}}, \quad (3.33)$$

mit

$$\mathbf{P}_{\mathbf{u}}(k) = \mathrm{E}\left\{\mathbf{u}(k)\mathbf{u}^{\mathrm{H}}(k)\right\} = \begin{bmatrix} \sigma_v^2(k) & 0 & 0 \\ 0 & \sigma_{w_1}^2(k) & 0 \\ 0 & 0 & \sigma_{w_2}^2(k) \end{bmatrix}. \quad (3.34)$$

Damit der Schätzwert optimal ist, muss der mittlere quadratische Fehler minimal sein. Dies ist gegeben, wenn der a-priori Schätzfehler $\mathbf{e}(k|k{-}1)$ orthogonal zu allen bis dato aufgetretenen Eingangsvektoren $\mathbf{y}(l)$ für $0 \leq l \leq k-1$ ist:

$$\begin{aligned} \mathbf{0} &= \mathrm{E}\left\{\mathbf{e}(k|k-1)\mathbf{y}^{\mathrm{H}}(l)\right\} \\ &= \mathrm{E}\left\{\left(\mathbf{A}(k|k-1)\mathbf{e}(k-1|k-1) + \mathbf{B}\mathbf{u}(k)\right)\mathbf{y}^{\mathrm{H}}(l)\right\} \\ &= \mathbf{A}(k|k-1)\underbrace{\mathrm{E}\left\{\mathbf{e}(k-1|k-1)\mathbf{y}^{\mathrm{H}}(l)\right\}}_{(1)} + \mathbf{B}\underbrace{\mathrm{E}\left\{\mathbf{u}(k)\mathbf{y}^{\mathrm{H}}(l)\right\}}_{(2)}. \end{aligned} \quad (3.35)$$

Der mit (1) abgekürzte Term in Gleichung 3.35 verschwindet, da $\hat{\mathbf{x}}(k-1|k-1)$ als optimaler Schätzwert angenommen wurde und somit für den zugehörigen Schätzfehler $\mathrm{E}\{\mathbf{e}(k–1|k–1)\mathbf{y}^{\mathrm{H}}(l)\} = \mathbf{0}$ für $0 \leq l \leq k–1$ gilt. Um zu zeigen, dass Ausdruck (2) Null ist, wird zuerst die Messgleichung 3.24 für $\mathbf{y}(l)$ eingesetzt und danach die Systemgleichung 3.23 für $\mathbf{x}(l)$:

$$\begin{aligned}
\mathrm{E}\left\{\mathbf{u}(k)\mathbf{y}^{\mathrm{H}}(l)\right\} &= \mathrm{E}\left\{\mathbf{u}(k)\left(\mathbf{C}^{\mathrm{H}}(l)\mathbf{x}(l)+\mathbf{z}(l)\right)^{\mathrm{H}}\right\} \\
&= \mathrm{E}\left\{\mathbf{u}(k)\left(\mathbf{C}^{\mathrm{H}}(l)\left(\mathbf{A}(l|l-1)\mathbf{x}(l-1)+\mathbf{B}\mathbf{u}(l)\right)+\mathbf{z}(l)\right)^{\mathrm{H}}\right\} \\
&= \mathrm{E}\left\{\mathbf{u}(k)\mathbf{x}^{\mathrm{H}}(l-1)\right\}\mathbf{A}^{\mathrm{H}}(l|l-1)\mathbf{C}(l) \\
&\quad +\mathrm{E}\left\{\mathbf{u}(k)\mathbf{u}^{\mathrm{H}}(l)\right\}\mathbf{B}^{\mathrm{T}}+\mathrm{E}\left\{\mathbf{u}(k)\mathbf{z}^{\mathrm{H}}(l)\right\}. \qquad (3.36)
\end{aligned}$$

Der mittlere Erwartungswert verschwindet, weil $0 \leq l \leq k-1$ gilt und $\mathbf{u}(k)$ weiß ist. Aufgrund der angenommenen Unkorreliertheit aller Rauschprozesse wird der hintere Term ebenfalls Null. Wie schon bei der Herleitung von $\mathbf{P}_{\mathbf{e}}(k|k–1)$ gezeigt, lässt sich $\mathbf{x}^{\mathrm{H}}(l–1)$ im vorderen Erwartungswert durch wiederholtes Einsetzen der Systemgleichung 3.23 in Gleichung 3.36 bis zum Anfangszustand $\mathbf{x}^{\mathrm{H}}(0)$ reduzieren. Dieser ist laut Annahme unkorreliert zu $\mathbf{u}(k)$ und somit identisch Null.

3.2.3 Korrektur

Im vorhergegangen Abschnitt wurde ein linearer, erwartungstreuer und optimaler a-priori Schätzwert $\hat{\mathbf{x}}(k|k–1)$ auf Grundlage aller Eingangsdaten bis zum Abtastzeitpunkt $k-1$ rekursiv berechnet. Dieser soll nun mit Hilfe des neue Eingangsvektors $\mathbf{y}(k)$ verbessert werden (a-posteriori Schätzwert). Als linearen, rekursiven Ansatz für diese a-posteriori Schätzung wird $\hat{\mathbf{x}}(k|k)$ gewählt [21]:

$$\hat{\mathbf{x}}(k|k) = \mathbf{\Psi}(k)\hat{\mathbf{x}}(k|k-1)+\mathbf{K}(k)\mathbf{y}(k), \qquad (3.37)$$

wobei die beiden Matrizen $\mathbf{\Psi}(k)$ und $\mathbf{K}(k)$ determiniert und noch zu bestimmen sind. Zunächst wird aber der a-posteriori Schätzfehler $\mathbf{e}(k|k)$ durch Einsetzen der Gleichungen 3.25, 3.28 und 3.37 in Gleichung 3.29 berechnet:

$$\begin{aligned}
\mathbf{e}(k) &= \mathbf{x}(k)-\mathbf{\Psi}(k)\hat{\mathbf{x}}(k|k-1)-\mathbf{K}(k)\mathbf{y}(k) \\
&= \mathbf{x}(k)-\mathbf{\Psi}(k)\left(\mathbf{x}(k)-\mathbf{e}(k|k-1)\right)-\mathbf{K}(k)\left(\mathbf{C}^{\mathrm{H}}(k)\mathbf{x}(k)+\mathbf{z}(k)\right) \\
&= \left(\mathbf{I}-\mathbf{\Psi}(k)-\mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\right)\mathbf{x}(k)+\mathbf{\Psi}(k)\mathbf{e}(k|k-1)-\mathbf{K}(k)\mathbf{z}(k). \qquad (3.38)
\end{aligned}$$

Für einen erwartungstreuen a-posteriori Schätzwert $\hat{\mathbf{x}}(k|k)$ ist es notwendig, dass dessen Schätzfehler aus Gleichung 3.38 mittelwertfrei ist:

$$\begin{aligned}
\mathbf{0} &= \mathrm{E}\{\mathbf{e}(k)\} \\
&= \mathrm{E}\left\{\left(\mathbf{I}-\mathbf{\Psi}(k)-\mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\right)\mathbf{x}(k)+\mathbf{\Psi}(k)\mathbf{e}(k|k-1)-\mathbf{K}(k)\mathbf{z}(k)\right\}. \qquad (3.39)
\end{aligned}$$

Da $\mathbf{e}(k|k-1)$ und $\mathbf{z}(k)$ mittelwertfrei sind, während $\mathbf{\Psi}(k)$ und $\mathbf{K}(k)$ als determiniert angenommen wurden, verschwinden die hinteren beiden Terme bei der Erwartungswertbildung. Eine einfache Möglichkeit, Mittelwertfreiheit zu erreichen, besteht jetzt darin, die Matrizen $\mathbf{\Psi}(k)$ und $\mathbf{K}(k)$ so zu wählen, dass der Klammerausdruck identisch Null wird. Daraus lässt sich eine Bedingung für $\mathbf{\Psi}(k)$ bestimmen:

$$\begin{aligned} \mathbf{0} &= \mathbf{I} - \mathbf{\Psi}(k) - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k) \\ \Rightarrow \mathbf{\Psi}(k) &= \mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k) \end{aligned} \tag{3.40}$$

Setzt man das Ergebnis in Gleichung 3.37 ein, erhält man nach wenigen Umformungen den gesuchten linearen und rekursiven Ansatz für einen erwartungstreuen a-posteriori Schätzwert $\hat{\mathbf{x}}(k|k)$:

$$\begin{aligned} \hat{\mathbf{x}}(k|k) &= \left(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\right)\hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)\mathbf{y}(k) \\ &= \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)\left(\mathbf{y}(k) - \mathbf{C}^{\mathrm{H}}(k)\hat{\mathbf{x}}(k|k-1)\right). \end{aligned} \tag{3.41}$$

Zur Verdeutlichung des Ergebnisses wird der vorhergesagte Messwert $\hat{\mathbf{y}}(k|k-1)$ definiert:

$$\hat{\mathbf{y}}(k|k-1) = \mathbf{C}^{\mathrm{H}}(k)\hat{\mathbf{x}}(k|k-1). \tag{3.42}$$

Dieser wird in Gleichung 3.41 eingesetzt, wodurch der a-posteriori Schätzwert $\hat{\mathbf{x}}(k|k)$ jetzt folgendermaßen geschrieben werden kann:

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)\left(\mathbf{y}(k) - \hat{\mathbf{y}}(k|k-1)\right). \tag{3.43}$$

$\hat{\mathbf{y}}(k|k-1)$ bezeichnet den auf Grundlage der vergangenen Eingangsvektoren $\mathbf{y}(l)$ für $0 \leq l \leq k-1$ vorhersagbaren Anteil von $\mathbf{y}(k)$. Der gesamte Klammerausdruck $\mathbf{y}(k) - \hat{\mathbf{y}}(k|k-1)$ stellt somit die *Innovation* durch den neuen Eingangsvektor $\mathbf{y}(k)$ dar, welche, durch eine Multiplikation mit der Matrix $\mathbf{K}(k)$ verstärkt, zu dem a-priori Schätzwert addiert wird. In diesem Zusammenhang bezeichnet man $\mathbf{K}(k)$ als *Kalman-Verstärkung* (englisch: *Kalman-Gain*).

Als letzte Anforderung an $\hat{\mathbf{x}}(k|k)$ verbleibt die Minimalität seines mittleren quadratischen Fehlers (optimaler Schätzwert). Dazu wird zunächst der a-posteriori Schätzfehler durch Einsetzten der Gleichungen 3.24, 3.42 und 3.43 in Gleichung 3.29 berechnet:

$$\begin{aligned} \mathbf{e}(k|k) &= \mathbf{x}(k) - \left(\hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)\left(\mathbf{y}(k) - \hat{\mathbf{y}}(k|k-1)\right)\right) \\ &= \mathbf{e}(k|k-1) - \mathbf{K}(k)\left(\mathbf{C}^{\mathrm{H}}(k)\mathbf{x}(k) + \mathbf{z}(k) - \mathbf{C}^{\mathrm{H}}(k)\hat{\mathbf{x}}(k|k-1)\right) \\ &= \mathbf{e}(k|k-1) - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\mathbf{e}(k|k-1) - \mathbf{K}(k)\mathbf{z}(k) \\ &= \left(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\right)\mathbf{e}(k|k-1) - \mathbf{K}(k)\mathbf{z}(k). \end{aligned} \tag{3.44}$$

Die Kalman-Verstärkung $\mathbf{K}(k)$ muss jetzt so gewählt werden, dass der mittlere quadratische Fehler $\overline{\mathbf{e}^2(k|k)} = \mathrm{E}\{\mathbf{e}^{\mathrm{H}}(k|k)\mathbf{e}(k|k)\}$ minimal wird. Anstatt $\overline{\mathbf{e}^2(k|k)}$

direkt nach $\mathbf{K}(k)$ zu differenzieren, wird der Weg über die Differentiation der Spur der Kovarianzmatrix $\mathbf{P_e}(k|k)$ gewählt:

$$\frac{\partial}{\partial \mathbf{K}(k)} \overline{e^2(k|k)} \;=\; \frac{\partial}{\partial \mathbf{K}(k)} \operatorname{tr}\left\{\mathbf{P_e}(k|k)\right\}. \tag{3.45}$$

Die Kovarianzmatrix $\mathbf{P_e}(k|k)$ berechnet sich durch Einsetzen von Gleichung 3.44 in Gleichung 3.31 zu:

$$\begin{aligned}
\mathbf{P_e}(k|k) &= \mathrm{E}\Big\{ \Big(\big(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\big)\, \mathbf{e}(k|k-1) - \mathbf{K}(k)\mathbf{z}(k) \Big) \cdots \\
&\quad \cdot \Big(\big(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\big)\, \mathbf{e}(k|k-1) - \mathbf{K}(k)\mathbf{z}(k) \Big)^{\mathrm{H}} \Big\} \\
&= \big(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\big)\, \mathbf{P_e}(k|k-1)\, \big(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\big)^{\mathrm{H}} \cdots \\
&\quad - \big(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\big) \underbrace{\mathrm{E}\left\{\mathbf{e}(k|k-1)\mathbf{z}^{\mathrm{H}}(k)\right\}}_{=\mathbf{0}} \mathbf{K}^{\mathrm{H}}(k) \cdots \\
&\quad - \mathbf{K}(k) \underbrace{\mathrm{E}\left\{\mathbf{z}(k)\mathbf{e}^{\mathrm{H}}(k|k-1)\right\}}_{=\mathbf{0}} \big(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\big)^{\mathrm{H}} \cdots \\
&\quad + \mathbf{K}(k)\mathbf{P_z}(k)\mathbf{K}^{\mathrm{H}}(k) \\
&= \big(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\big)\, \mathbf{P_e}(k|k-1)\, \big(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\big)^{\mathrm{H}} \cdots \\
&\quad + \mathbf{K}(k)\mathbf{P_z}(k)\mathbf{K}^{\mathrm{H}}(k),
\end{aligned} \tag{3.46}$$

wobei $\mathbf{P_z}(k)$ die Kovarianzmatrix des zusätzlichen Messrauschens bezeichnet und gemäß Gleichung 3.76 für $N = 2$ definiert ist. Die Kreuzkorrelationsterme verschwinden, weil der a-priori Schätzfehler auf Grundlage der Daten bis zum Zeitpunkt k–1 und das Messrauschen zum Zeitpunkt k unkorreliert sind (siehe Prädiktion in Abschnitt 3.2.2). Die mathematischen Zusammenhänge der nachfolgenden Vektor/Matrix-Ableitung sind in [9, 10, 12, 35, 36] beschrieben.

Um das optimale $\mathbf{K}(k)$ zu bestimmen, werden nun der Zusammenhang aus Gleichung 3.45 und die Kettenregel auf Gleichung 3.46 angewendet:

$$\begin{aligned}
\mathbf{0} &= \frac{\partial}{\partial \mathbf{K}(k)} \operatorname{tr}\Big\{ \big(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\big)\, \mathbf{P_e}(k|k-1)\, \big(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\big)^{\mathrm{H}} \cdots \\
&\quad + \mathbf{K}(k)\mathbf{P_z}(k)\mathbf{K}^{\mathrm{H}}(k) \Big\} \\
&= 2\,\big(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\big)\, \mathbf{P_e}(k|k-1) \frac{\partial}{\partial \mathbf{K}(k)} \operatorname{tr}\Big\{ \big(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\big) \Big\} \cdots \\
&\quad + 2\mathbf{K}(k)\mathbf{P_z}(k) \\
&= \big(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\big)\, \mathbf{P_e}(k|k-1)\, (-\mathbf{C}(k)) + \mathbf{K}(k)\mathbf{P_z}(k) \\
&= \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\mathbf{P_e}(k|k-1)\mathbf{C}(k) - \mathbf{P_e}(k|k-1)\mathbf{C}(k) + \mathbf{K}(k)\mathbf{P_z}(k).
\end{aligned} \tag{3.47}$$

Unter der Voraussetzung, dass die Inverse $(\mathbf{C}^{\mathrm{H}}(k)\mathbf{P}_{\mathbf{e}}(k|k-1)\mathbf{C}(k) + \mathbf{P}_{\mathbf{z}}(k))^{-1}$ existiert, kann Gleichung 3.47 nach der optimalen Kalman-Verstärkung $\mathbf{K}(k)$ aufgelöst werden:

$$\mathbf{K}(k) \;=\; \mathbf{P}_{\mathbf{e}}(k|k-1)\mathbf{C}(k)\left(\mathbf{C}^{\mathrm{H}}(k)\mathbf{P}_{\mathbf{e}}(k|k-1)\mathbf{C}(k)+\mathbf{P}_{\mathbf{z}}(k)\right)^{-1}. \tag{3.48}$$

Die Formel 3.46 für die Kovarianzmatrix des a-posteriori Fehlers kann weiter vereinfacht werden. Dazu wird zunächst die rechte Seite der Gleichung ausmultipliziert:

$$\begin{aligned}\mathbf{P}_{\mathbf{e}}(k|k) \;&=\; \left(\mathbf{P}_{\mathbf{e}}(k|k-1) - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\mathbf{P}_{\mathbf{e}}(k|k-1)\right)\left(\mathbf{I}-\mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\right)^{\mathrm{H}}\cdots\\ &\quad +\mathbf{K}(k)\mathbf{P}_{\mathbf{z}}(k)\mathbf{K}^{\mathrm{H}}(k)\\ &=\; \mathbf{P}_{\mathbf{e}}(k|k-1) - \mathbf{P}_{\mathbf{e}}(k|k-1)\mathbf{C}(k)\mathbf{K}^{\mathrm{H}}(k) - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\mathbf{P}_{\mathbf{e}}(k|k-1)\cdots\\ &\quad +\mathbf{K}(k)\left(\mathbf{C}^{\mathrm{H}}(k)\mathbf{P}_{\mathbf{e}}(k|k-1)\mathbf{C}(k)+\mathbf{P}_{\mathbf{z}}(k)\right)\mathbf{K}^{\mathrm{H}}(k).\end{aligned}$$

Wird jetzt in dem quadratischen Term der letzten Zeile für $\mathbf{K}(k)$, jedoch nicht für $\mathbf{K}^{\mathrm{H}}(k)$, die Gleichung 3.48 für die optimale Kalman-Verstärkung eingesetzt, vereinfacht sich die Gleichung schließlich zu:

$$\begin{aligned}\mathbf{P}_{\mathbf{e}}(k|k) &= \mathbf{P}_{\mathbf{e}}(k|k-1) - \mathbf{P}_{\mathbf{e}}(k|k-1)\mathbf{C}(k)\mathbf{K}^{\mathrm{H}}(k) - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\mathbf{P}_{\mathbf{e}}(k|k-1)\cdots\\ &\quad +\left(\mathbf{P}_{\mathbf{e}}(k|k-1)\mathbf{C}(k)\left(\mathbf{C}^{\mathrm{H}}(k)\mathbf{P}_{\mathbf{e}}(k|k-1)\mathbf{C}(k)+\mathbf{P}_{\mathbf{z}}(k)\right)^{-1}\right)\cdots\\ &\quad \left(\mathbf{C}^{\mathrm{H}}(k)\mathbf{P}_{\mathbf{e}}(k|k-1)\mathbf{C}(k)+\mathbf{P}_{\mathbf{z}}(k)\right)\mathbf{K}^{\mathrm{H}}(k)\\ &= \mathbf{P}_{\mathbf{e}}(k|k-1) - \mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\mathbf{P}_{\mathbf{e}}(k|k-1)\\ &= \left(\mathbf{I}-\mathbf{K}(k)\mathbf{C}^{\mathrm{H}}(k)\right)\mathbf{P}_{\mathbf{e}}(k|k-1).\end{aligned} \tag{3.49}$$

In Abbildung 3.3 ist der hergeleitete Kalman-Filter Algorithmus für $N = 2$ Mikrofone als Signalflussdiagramm visualisiert. Wiederum sind aus Gründen der Übersichtlichkeit alle Größen ohne den Zeitindex k dargestellt. Vergleicht man dieses Diagramm mit dem aus Abbildung 3.2, so ist deutlich zu erkennen, dass das Kalman-Filter das zugrunde liegende Signalmodell nachbildet. Zur Verdeutlichung ist der zur Messmatrix $\mathbf{C}(k)$ korrespondierende Teil mit einem gestrichelten Kasten eingerahmt.

Damit ist die Herleitung des rekursiven Teils des Verfahrens abgeschlossen. Im nächsten Abschnitt wird nun die Initialisierung des Algorithmus' erläutert.

3.2.4 Initialisierung

Zum Abtastzeitpunkt $k = 0$ liegen aufgrund der eingangs getroffen Annahmen neben dem aktuellen Messwert $\mathbf{y}(0)$ keine weiteren Daten bzw. vergangene Schätzwerte vor. Daher wird für den ersten Rechenschritt des Algorithmus' folgender, wiederum linearer Ansatz gewählt [19]:

$$\hat{\mathbf{x}}(0|0) \;=\; \mathbf{K}(0)\mathbf{y}(0). \tag{3.50}$$

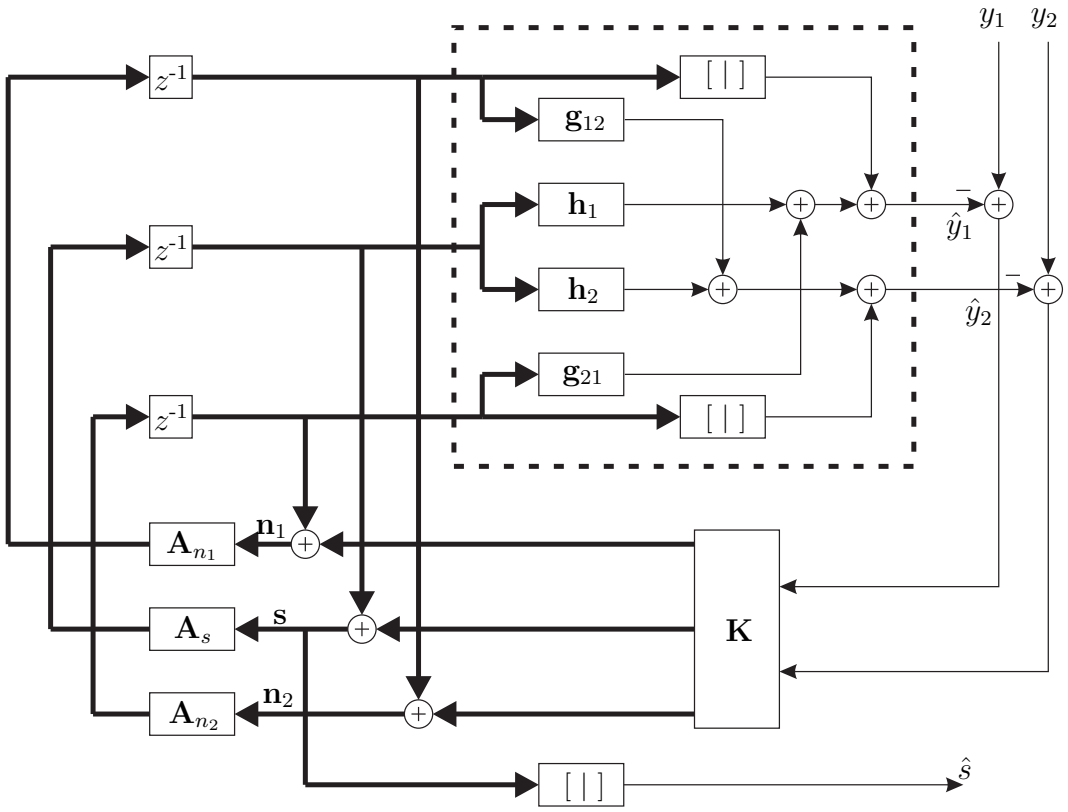

Abbildung 3.3: *Kalman-Filter in Zustandsraumdarstellung.*

Setzt man für $\mathbf{y}(0)$ die Messgleichung 3.25 ein und beachtet, dass Mittelwertfreiheit für $\mathbf{x}(0)$ und $\mathbf{z}(k)$ angenommen wurde, so lässt sich zeigen, dass $\hat{\mathbf{x}}(0|0)$ erwartungstreu ist:

$$\begin{aligned} \mathrm{E}\left\{\hat{\mathbf{x}}(0|0)\right\} &= \mathrm{E}\left\{\mathbf{K}(0)\left(\mathbf{C}^{\mathrm{H}}(0)\mathbf{x}(0)+\mathbf{z}(0)\right)\right\} \\ &= \mathbf{K}(0)\mathbf{C}^{\mathrm{H}}(0)\mathrm{E}\left\{\mathbf{x}(0)\right\}+\mathbf{K}(0)\mathrm{E}\left\{\mathbf{z}(0)\right\}=\mathbf{0}. \end{aligned} \tag{3.51}$$

Um einen optimalen Schätzwert $\hat{\mathbf{x}}(0|0)$ zu erhalten, werden wiederum der mittlere quadratische Fehler minimiert und daraus das optimale $\mathbf{K}(0)$ berechnet. Hierzu werden in den Fehlervektor $\mathbf{e}(0|0)$ nacheinander die Gleichungen 3.50 und 3.25 eingesetzt:

$$\begin{aligned} \mathbf{e}(0|0) &= \mathbf{x}(0)-\hat{\mathbf{x}}(0|0)=\mathbf{x}(0)-\mathbf{K}(0)\mathbf{y}(0) \\ &= \mathbf{x}(0)-\mathbf{K}(0)\mathbf{C}^{\mathrm{H}}(0)\mathbf{x}(0)-\mathbf{K}(0)\mathbf{z}(0) \\ &= \left(\mathbf{I}-\mathbf{K}(0)\mathbf{C}^{\mathrm{H}}(0)\right)\mathbf{x}(0)-\mathbf{K}(0)\mathbf{z}(0). \end{aligned} \tag{3.52}$$

Daraus lässt sich mit Hilfe von Gleichung 3.31 die zugehörige Kovarianzmatrix $\mathbf{P_e}(0|0)$ berechnen:

$$\begin{aligned}
\mathbf{P_e}(0|0) &= \mathrm{E}\Big\{ \left(\left(\mathbf{I} - \mathbf{K}(0)\mathbf{C}^{\mathrm{H}}(0)\right)\mathbf{x}(0) - \mathbf{K}(0)\mathbf{z}(0)\right) \cdots \\
&\quad \left(\left(\mathbf{I} - \mathbf{K}(0)\mathbf{C}^{\mathrm{H}}(0)\right)\mathbf{x}(0) - \mathbf{K}(0)\mathbf{z}(0)\right)^{\mathrm{H}} \Big\} \\
&= \left(\mathbf{I} - \mathbf{K}(0)\mathbf{C}^{\mathrm{H}}(0)\right) \mathrm{E}\left\{\mathbf{x}(0)\mathbf{x}^{\mathrm{H}}(0)\right\} \left(\mathbf{I} - \mathbf{K}(0)\mathbf{C}^{\mathrm{H}}(0)\right)^{\mathrm{H}} \cdots \\
&\quad - \left(\mathbf{I} - \mathbf{K}(0)\mathbf{C}^{\mathrm{H}}(0)\right) \mathrm{E}\left\{\mathbf{x}(0)\mathbf{z}^{\mathrm{H}}(0)\right\} \mathbf{K}^{\mathrm{H}}(0) \cdots \\
&\quad - \mathbf{K}(0)\mathrm{E}\left\{\mathbf{z}(0)\mathbf{x}^{\mathrm{H}}(0)\right\} \left(\mathbf{I} - \mathbf{K}(0)\mathbf{C}^{\mathrm{H}}(0)\right)^{\mathrm{H}} + \mathbf{K}(0)\mathbf{P_z}(0)\mathbf{K}^{\mathrm{H}}(0) \\
&= \left(\mathbf{I} - \mathbf{K}(0)\mathbf{C}^{\mathrm{H}}(0)\right) \mathbf{P_x}(0) \left(\mathbf{I} - \mathbf{K}(0)\mathbf{C}^{\mathrm{H}}(0)\right)^{\mathrm{H}} + \mathbf{K}(0)\mathbf{P_z}(0)\mathbf{K}^{\mathrm{H}}(0), \quad (3.53)
\end{aligned}$$

wobei $\mathbf{P_x}(0)$ die als bekannt angenommen Kovarianzmatrix des Anfangszustands $\mathbf{x}(0)$ bezeichnet.

Analog zum Vorgehen in Gleichung 3.47 wird nun das optimale $\mathbf{K}(0)$ durch Differentiation der Spur von $\mathbf{P_e}(0|0)$ bestimmt:

$$\begin{aligned}
\mathbf{0} &= \frac{\partial}{\partial \mathbf{K}(0)} \mathrm{tr} \Big\{ \left(\mathbf{I} - \mathbf{K}(0)\mathbf{C}^{\mathrm{H}}(0)\right) \mathbf{P_x}(0) \left(\mathbf{I} - \mathbf{K}(0)\mathbf{C}^{\mathrm{H}}(0)\right)^{\mathrm{H}} \\
&\quad + \mathbf{K}(0)\mathbf{P_z}(0)\mathbf{K}^{\mathrm{H}}(0) \Big\} \\
&= \mathbf{K}(0)\mathbf{C}^{\mathrm{H}}(0)\mathbf{P_x}(0)\mathbf{C}(0) - \mathbf{P_x}(0)\mathbf{C}(0) + \mathbf{K}(0)\mathbf{P_z}(0). \quad (3.54)
\end{aligned}$$

Unter der Bedingung, dass die Inverse existiert, kann man nach $\mathbf{K}(0)$ freistellen und erhält:

$$\mathbf{K}(0) = \mathbf{P_x}(0)\mathbf{C}(0) \left(\mathbf{C}^{\mathrm{H}}(0)\mathbf{P_x}(0)\mathbf{C}(0) + \mathbf{P_z}(0)\right)^{-1}. \quad (3.55)$$

Durch Ausmultiplizieren von Gleichung 3.53 und Einsetzen von Gleichung 3.55 vereinfacht sich die Formel für die Kovarianzmatrix $\mathbf{P_e}(0|0)$ schließlich zu:

$$\mathbf{P_e}(0|0) = \left(\mathbf{I} - \mathbf{K}(0)\mathbf{C}^{\mathrm{H}}(0)\right) \mathbf{P_x}(0). \quad (3.56)$$

In der Anwendung initialisiert man die Komponenten von $\mathbf{P_x}(0)$, die nicht automatisch Null sind, mit der Leistung der Mikrofonsignale für $k = 0$ [19]. Die Wahl von $\mathbf{P_z}(k)$ wird in Abschnitt 3.4.2 behandelt.

Eine Zusammenfassung der hergeleiteten Gleichungen des Kalman-Filters ist in Tabelle 3.1 abgebildet.

3.3 Erweiterung auf mehr als zwei Kanäle

Zu Beginn der Herleitung wurden verschiedene Annahmen getroffen, von denen eine die Beschränkung auf zwei Mikrofone war. Diese Einschränkung wird im

Tabelle 3.1: *Zusammenfassung der Gleichungen für das Kalman-Filter-Verfahren.*

Initialisierung
$\mathbf{K}(0) = \mathbf{P}_\mathbf{x}(0)\mathbf{C}(0)\left(\mathbf{C}^\mathrm{H}(0)\mathbf{P}_\mathbf{x}(0)\mathbf{C}(0) + \mathbf{P}_\mathbf{z}(0)\right)^{-1}$
$\hat{\mathbf{x}}(0\|0) = \mathbf{K}(0)\mathbf{y}(0)$
$\mathbf{P}_\mathbf{e}(0\|0) = \left(\mathbf{I} - \mathbf{K}(0)\mathbf{C}^\mathrm{H}(0)\right)\mathbf{P}_\mathbf{x}(0)$
Prädiktion
$\hat{\mathbf{x}}(k\|k-1) = \mathbf{A}(k\|k-1)\hat{\mathbf{x}}(k-1\|k-1)$
$\mathbf{P}_\mathbf{e}(k\|k-1) = \mathbf{A}(k\|k-1)\mathbf{P}_\mathbf{e}(k-1\|k-1)\mathbf{A}^\mathrm{H}(k\|k-1) + \mathbf{B}\mathbf{P}_\mathbf{u}(k)\mathbf{B}^\mathrm{T}$
Korrektur
$\mathbf{K}(k) = \mathbf{P}_\mathbf{e}(k\|k-1)\mathbf{C}(k)\left(\mathbf{C}^\mathrm{H}(k)\mathbf{P}_\mathbf{e}(k\|k-1)\mathbf{C}(k) + \mathbf{P}_\mathbf{z}(k)\right)^{-1}$
$\hat{\mathbf{x}}(k\|k) = \hat{\mathbf{x}}(k\|k-1) + \mathbf{K}(k)\left(\mathbf{y}(k) - \mathbf{C}^\mathrm{H}(k)\hat{\mathbf{x}}(k\|k-1)\right)$
$\mathbf{P}_\mathbf{e}(k\|k) = \left(\mathbf{I} - \mathbf{K}(k)\mathbf{C}^\mathrm{H}(k)\right)\mathbf{P}_\mathbf{e}(k\|k-1)$

nun folgenden Abschnitt fallen gelassen und die Erweiterung des hergeleiteten Verfahrens auf theoretisch beliebig viele Kanäle vorgestellt. Alle anderen Annahmen, insbesondere die, dass stets nur ein Sprecher aktiv ist, bleiben bestehen. Im letzten Teil des Kapitels wird der Einfluss dieser Erweiterung anhand des durch sie verursachten numerischen Aufwands diskutiert.

Das Kalman-Filter liefert einen Schätzwert für die ungestörte Sprachkomponente ausgehend von den gestört vorliegenden Signalen an den Mikrofonausgängen. Erhöht man die Anzahl der Mikrofone, erhöht sich ebenfalls die Anzahl der Eingangssignale des Algorithmus', während unabhängig davon stets nur eine Sprachquelle im Signalmodell existiert. Es stellt sich also die Frage, ob eine große Anzahl von Eingangssignalen gegenüber einer kleinen Vorteile bei der Schätzung dieses Quellensignals bewirkt. Dies wird im Kapitel 6 diskutiert.

Ein Nachteil einer Erhöhung der Mikrofonanzahl ist, dass die Dimension der verwendeten Matrizen und Vektoren ebenfalls mit steigender Anzahl von Kanälen zunimmt, was einen höheren numerischen Aufwand bedeutet. Dies wirkt sich besonders dort aus, wo der Rechenaufwand schneller als linear mit der Dimension der Matrizen wächst. Beispielsweise muss für die Berechnung der Kalman-Verstärkung eine Matrix invertiert werden. Der numerische Aufwand hierfür steigt kubisch mit der Dimension der zu invertierenden Matrix [8], da diese keine besondere Struktur aufweist. Hinzu kommt, dass jede Rauschquelle ebenfalls modelliert wird. Deren Anzahl bleibt bei steigender Mikrofonanzahl allerdings nicht konstant, sondern nimmt gleichermaßen zu.

Die gleiche Argumentation kann für die Schätzung der Parameter, die für den Betrieb des Kalman-Filters notwendig sind, angewendet werden. Dies wird zusammen mit den Parameter-Schätzverfahren im nächsten Kapitel behandelt. Im

Folgenden werden daher nur die Auswirkungen auf das in Abschnitt 3.1 vorgestellte Signalmodell und das in Abschnitt 3.2 hergeleitete MISO-Kalman-Filter diskutiert. Dabei wird sich hauptsächlich auf formelmäßige Zusammenhänge beschränkt, da die grafische Darstellung durch Signalflussdiagramme mit steigender Komplexität sehr unübersichtlich wird. Für die nun folgenden Überlegungen wird eine Mikrofonanzahl $N > 2$ angenommen.

Erweiterung von System- und Messgleichung

Für die Erweiterung der Systemgleichung 3.23 werden zunächst die Vektoren aus Gleichung 3.18 erweitert:

$$\mathbf{x}(k) = \begin{bmatrix} \mathbf{s}(k) \\ \mathbf{n}_1(k) \\ \mathbf{n}_2(k) \\ \vdots \\ \mathbf{n}_N(k) \end{bmatrix}, \qquad \mathbf{u}(k) = \begin{bmatrix} v(k) \\ w_1(k) \\ w_2(k) \\ \vdots \\ w_N(k) \end{bmatrix}. \tag{3.57}$$

Entsprechend wird nun mit den Matrizen $\mathbf{A}(k|k-1)$ und $\mathbf{B}$ verfahren. Diese werden entlang ihrer Hauptdiagonalen erweitert. Am Beispiel der $\mathbf{A}$-Matrix erhält man:

$$\mathbf{A}(k|k-1) = \begin{bmatrix} \mathbf{A}_s(k|k-1) & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{n_1}(k|k-1) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{n_2}(k|k-1) & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \cdots & \mathbf{A}_{n_N}(k|k-1) \end{bmatrix}. \tag{3.58}$$

Aufgrund der anfangs getroffenen Annahmen ist der oben dargestellte Weg über die Erweiterung der Zustands- und Anregungsvektoren mit anschließender Anpassung der Matrizen der einzig mögliche. Für die Messgleichung gilt dies nicht. Zwar ergeben sich auch hier einige erweiterte Größen automatisch, zum Beispiel alle Vektoren, die Struktur der erweiterten Messgleichung $\mathbf{C}(k)$ weist allerdings Freiheitsgrade auf.

Zum einen ist eine Erweiterung gemäß dem gewählten Signalmodell denkbar. Das heißt, es gibt jeweils ein Kreuzfilter für jede mögliche Kombination von Eingangs-

kanalpaaren:

$$\mathbf{C}(k) = \begin{bmatrix} \mathbf{h}_1^{\mathrm{H}}(k) & [\,|\,]^{\mathrm{T}} & \mathbf{g}_{21}^{\mathrm{H}}(k) & \mathbf{g}_{31}^{\mathrm{H}}(k) & \cdots & \mathbf{g}_{N1}^{\mathrm{H}}(k) \\ \mathbf{h}_2^{\mathrm{H}}(k) & \mathbf{g}_{12}^{\mathrm{H}}(k) & [\,|\,]^{\mathrm{T}} & \mathbf{g}_{32}^{\mathrm{H}}(k) & \cdots & \mathbf{g}_{N2}^{\mathrm{H}}(k) \\ \mathbf{h}_3^{\mathrm{H}}(k) & \mathbf{g}_{13}^{\mathrm{H}}(k) & \mathbf{g}_{23}^{\mathrm{H}}(k) & [\,|\,]^{\mathrm{T}} & \cdots & \mathbf{g}_{N3}^{\mathrm{H}}(k) \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}_N^{\mathrm{H}}(k) & \mathbf{g}_{1N}^{\mathrm{H}}(k) & \mathbf{g}_{2N}^{\mathrm{H}}(k) & \mathbf{g}_{3N}^{\mathrm{H}}(k) & \cdots & [\,|\,]^{\mathrm{T}} \end{bmatrix}. \quad (3.59)$$

Zum anderen können beispielsweise auch nur die Kreuzkorrelationen bestimmter Kanalpaare berücksichtigt werden, während die übrigen vernachlässigt werden. Diese kann der Entwickler frei wählen. Zudem müssen die Impulsantworten dieser vernachlässigten Kreuzfilter nicht mehr geschätzt werden. Ein mögliches Beispiel ist, nur die Kreuzkorrelationsfunktionen räumlich direkt benachbarter Kanäle zu verwenden. Sind die Mikrofone zum Beispiel in einer linearen Array-Struktur (siehe Abschnitt 2.4) angeordnet, so ist in diesem Fall die zu erwartende Kreuzkorrelation zwischen zwei direkt benachbarten Mikrofonen größer als zwischen zwei nicht direkt benachbarten. Für eine solche lineare Mikrofonanordnung kann demnach beispielsweise folgende Darstellung verwendet werden:

$$\mathbf{C}(k) = \begin{bmatrix} \mathbf{h}_1^{\mathrm{H}}(k) & [\,|\,]^{\mathrm{T}} & \mathbf{g}_{21}^{\mathrm{H}}(k) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{h}_2^{\mathrm{H}}(k) & \mathbf{g}_{12}^{\mathrm{H}}(k) & [\,|\,]^{\mathrm{T}} & \mathbf{g}_{32}^{\mathrm{H}}(k) & \cdots & \mathbf{0} \\ \mathbf{h}_3^{\mathrm{H}}(k) & \mathbf{0} & \mathbf{g}_{23}^{\mathrm{H}}(k) & [\,|\,]^{\mathrm{T}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}_N^{\mathrm{H}}(k) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & [\,|\,]^{\mathrm{T}} \end{bmatrix}. \quad (3.60)$$

Die Darstellung in Gleichung 3.60 vernachlässigt alle Kreuzfilterpaare, welche nicht von zwei benachbarten Mikrofonen herrühren. Sie soll aber lediglich als Beispiel für die vorhandenen Möglichkeiten dienen und nicht eine Beschränkung auf diese besondere Anordnungen suggerieren. Bei einer wahllosen Anordnung der Mikrofone im Fahrzeuginnenraum würde sich mit der gleichen physikalischen Argumentation eine entsprechend weniger strukturierte Messmatrix ergeben, in der ebenfalls nur die Kreuzkorrelationen zwischen Mikrofonen mit paarweise minimalem Abstand berücksichtigt werden.

Wie sich bei der Diskussion der Simulationsergebnisse in Kapitel 6 zeigen wird, spielt die Wahl zwischen der Matrixstruktur aus Gleichung 3.59 und der aus Gleichung 3.60 eine untergeordnete Rolle bezüglich der erreichbaren Geräuschreduktion und Sprachqualität des Verfahrens.

Reduktion auf einen Kanal

Der Vollständigkeit halber sei abschließend erwähnt, dass mit den analogen Überlegungen die Mikrofonanzahl auch auf Eins reduziert werden kann. In diesem

Fall verschwinden alle Kreuzfilter, die Matrixinversion bei der Berechnung der Kalman-Verstärkung wird zu einer skalaren Division und man erhält ein einkanaliges Verfahren, welches dem nach [38] ähnelt.

Allerdings unterscheidet sich dieses von dem hier vorgestellten Algorithmus in zwei Punkten. Für $N = 1$ ergibt sich für das Signalmodell aus Abschnitt 3.1 folgende Messgleichung:

$$y(k) \quad = \quad \mathbf{h}^{\mathrm{H}}(k)\mathbf{x}(k) + z(k). \tag{3.61}$$

Ein Vergleich mit [38] ergibt folgende Unterschiede:

- Die Matrix $\mathbf{C}(k)$ reduziert sich für $N = 1$ auf eine Raumimpulsantwort $\mathbf{h}(k)$ und nicht auf den Ausschneidevektor [|] entsprechender Dimension. Dadurch ist das hier vorgestellte Verfahren in der Lage unter der Annahme einer hinreichend guten Schätzung von $\mathbf{h}(k)$ neben der Geräuschreduktion eine zusätzliche Enthallung zu bewirken.
- Die Aufnahme des zusätzlichen weißen Messrauschens $z(k)$ ermöglicht bei entsprechender Wahl von $\sigma_z^2(k)$ die Gewichtung des Messwerts $y(k)$ innerhalb des Kalman-Algorithmus' zu beeinflussen (siehe Abschnitt 3.4.2).

3.4 Analyse der Kalman-Filtergleichungen

In diesem Abschnitt werden die hergeleiteten Gleichungen des Kalman-Filters zur Veranschaulichung dessen Funktionsweise analysiert und interpretiert. Dabei wird ein besonderer Augenmerk auf die Bedeutung der Messmatrix $\mathbf{C}(k)$ gelegt. Zusätzlich werden mögliche Verbesserungen der Robustheit des Verfahrens, die durch Veränderungen im Algorithmusablauf erreicht werden können, aufgezeigt[3].

Durchgriffsfreiheit

Eine Besonderheit des hier gewählten Signalmodells ist, dass es nicht *durchgriffsfrei* ist. Setzt man die Systemgleichung 3.23 in die Messgleichung 3.24 ein, erhält man:

$$\mathbf{y}(k) \quad = \quad \mathbf{C}^{\mathrm{H}}(k)\left(\mathbf{A}(k|k-1)\mathbf{x}(k-1) + \mathbf{B}\mathbf{u}(k)\right). \tag{3.62}$$

Da $\mathbf{u}(k)$ ein Eingang ist und die gleiche Zeitabhängigkeit wie $\mathbf{y}(k)$ hat, erscheint eine Anregung am Eingang unverzögert am Ausgang [19]. Es ist also ein *Durch-*

[3] Eine Analyse der Funktionsweise sowohl aus Sichtweise der Signalverarbeitung als auch aus der Regelungstechnik findet sich in [17].

griff des Eingangs auf den Ausgang vorhanden. Eine solche direkte Kopplung erlaubt schnelle Systemreaktionen[4], birgt aber die Gefahr eines Aufschwingens.

3.4.1 Einfluss der Messmatrix

Für die Analyse der Messmatrix wird das Pseudo-Messrauschen vernachlässigt ($\mathbf{z}(k) = \mathbf{0}$), da es nur zur Steuerung der Robustheit verwendet wird und im Signalmodell aus Abschnitt 3.1 nicht vorkommt. Dies erleichtert die anschließenden Umformungen und schränkt die Allgemeinheit der Schlussfolgerungen nicht ein.

Ausgangspunkt ist die Gleichung 3.48 für die Kalman-Verstärkung mit $\mathbf{P_z}(k) = \mathbf{0}$:

$$\mathbf{K}(k) \;=\; \mathbf{P_e}(k|k-1)\mathbf{C}(k)\left(\mathbf{C}^{\mathrm{H}}(k)\mathbf{P_e}(k|k-1)\mathbf{C}(k)\right)^{-1}, \tag{3.63}$$

welche von links mit $\mathbf{C}^{\mathrm{H}}(k)$ multipliziert wird:

$$\mathbf{C}^{\mathrm{H}}(k)\mathbf{K}(k) \;=\; \mathbf{C}^{\mathrm{H}}(k)\mathbf{P_e}(k|k-1)\mathbf{C}(k)\left(\mathbf{C}^{\mathrm{H}}(k)\mathbf{P_e}(k|k-1)\mathbf{C}(k)\right)^{-1} = \mathbf{I} \tag{3.64}$$

Ein weiteres Multiplizieren mit $\mathbf{C}(k)$ und anschließender Inversion der Matrix $\mathbf{C}(k)\mathbf{C}^{\mathrm{H}}(k)$ führt nicht zum Ziel, weil letztere nicht invertierbar ist. Wie oben beschrieben weist $\mathbf{C}(k)$ mehr Zeilen (mindestens $N+1$) als Spalten (N) auf. Für ihren Rang gilt daher:

$$\mathrm{rank}\{\mathbf{C}(k)\} \leq N. \tag{3.65}$$

Da der Rang einer Matrix durch eine Multiplikation nicht größer werden kann [8], gilt für den Rang des $(N+1)\times(N+1)$ großen Produktes $\mathbf{C}(k)\mathbf{C}^{\mathrm{H}}(k)$ ebenfalls:

$$\mathrm{rank}\{\mathbf{C}(k)\mathbf{C}^{\mathrm{H}}(k)\} \leq N. \tag{3.66}$$

Das Matrixprodukt weist nicht vollen Spaltenrang auf und kann daher nicht invertiert werden.

Um dennoch weiterrechnen zu können, werden die mathematischen Hilfsmittel *Verallgemeinerte Inverse* und deren Spezialfall *Pseudoinverse* verwendet [34, 42]. Dazu werden folgende vier Bedingungen betrachtet:

$$(1) \quad \mathbf{C}(k)\mathbf{C}'(k)\mathbf{C}(k) = \mathbf{C}(k), \tag{3.67}$$

$$(2) \quad \mathbf{C}'(k)\mathbf{C}(k)\mathbf{C}'(k) = \mathbf{C}'(k), \tag{3.68}$$

$$(3) \quad \mathbf{C}(k)\mathbf{C}'(k) \text{ ist symmetrisch}, \tag{3.69}$$

$$(4) \quad \mathbf{C}'(k)\mathbf{C}(k) \text{ ist symmetrisch}. \tag{3.70}$$

[4] Ein solches System wird als *sprungfähig* bezeichnet.

Eine Matrix $\mathbf{C}'(k)$, die die erste Gleichung erfüllt, mindestens eine der verbleibenden aber nicht, heißt verallgemeinerte Inverse und wird mit $\mathbf{C}'(k) = \mathbf{C}^{-}(k)$ notiert. Sie existiert zwar immer, ist aber keine eindeutige Lösung eines Gleichungssystems.

Unter allen möglichen Lösungen existiert aber stets genau eine, welche alle vier Gleichungen erfüllt. Dieser als Pseudoinverse benannte und mit $\mathbf{C}'(k) = \mathbf{C}^{+}(k)$ bezeichnete Spezialfall resultiert als einzige aller möglichen Lösungen in einer minimalen L_2-Norm der Lösung.

Unter der Annahme, dass die Messmatrix $\mathbf{C}(k)$ vollen Spaltenrang aufweist[5], das heißt $\text{rank}\{\mathbf{C}(k)\} = N$, kann $\mathbf{C}'(k)$ wie folgt direkt berechnet werden [36]:

$$\mathbf{C}'(k) \;=\; \mathbf{C}(k)\left(\mathbf{C}^{\mathrm{H}}(k)\mathbf{C}(k)\right)^{-1}. \tag{3.71}$$

Damit lässt sich Gleichung 3.64 auflösen und man erhält:

$$\mathbf{K}(k) \;=\; \mathbf{C}'(k)\mathbf{I} = \mathbf{C}(k)\left(\mathbf{C}^{\mathrm{H}}(k)\mathbf{C}(k)\right)^{-1}. \tag{3.72}$$

Eingesetzt in Gleichung 3.41 erhält man diesmal (wiederum nach Einsetzen der Messgleichung):

$$\begin{aligned}
\hat{\mathbf{x}}(k|k) &= \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)\left(\mathbf{y}(k) - \mathbf{C}^{\mathrm{H}}(k)\hat{\mathbf{x}}(k|k-1)\right) \\
&= \hat{\mathbf{x}}(k|k-1) + \mathbf{C}'(k)\left(\mathbf{y}(k) - \mathbf{C}^{\mathrm{H}}(k)\hat{\mathbf{x}}(k|k-1)\right) \\
&= \hat{\mathbf{x}}(k|k-1) + \mathbf{C}(k)\left(\mathbf{C}^{\mathrm{H}}(k)\mathbf{C}(k)\right)^{-1}\left(\mathbf{C}^{\mathrm{H}}(k)\mathbf{x}(k) - \mathbf{C}^{\mathrm{H}}(k)\hat{\mathbf{x}}(k|k-1)\right) \\
&= \hat{\mathbf{x}}(k|k-1) + \mathbf{C}(k)\left(\mathbf{C}^{\mathrm{H}}(k)\mathbf{C}(k)\right)^{-1}\mathbf{C}^{\mathrm{H}}(k)\mathbf{e}(k|k-1).
\end{aligned} \tag{3.73}$$

In diesem Fall kann der Zustandsvektor $\mathbf{x}(k)$ nicht mehr direkt aus dem Mikrofonsignal gemessen werden. Dies liegt daran, dass die Kalman-Verstärkung jetzt den N-dimensionalen Mikrofonsignalvektor auf den $p+Nq > N$-dimensionalen Zustandsvektor abbildet. Es liegen also mehr Zustände als Eingänge vor, was keiner eindeutigen Abbildung entspricht.

Ein Vergleich von Gleichung 3.63 mit Gleichung 3.72 zeigt, dass die Matrix $\mathbf{C}'(k)$ den Spezialfall der Kalman-Verstärkung für $\mathbf{P}_{\mathbf{e}}(k|k-1) = \mathbf{I}$ darstellt. Eine a-priori Kovarianzmatrix der Form:

$$\mathbf{P}_{\mathbf{e}}(k|k-1) \;=\; \beta\mathbf{I} \qquad \text{mit } \beta \in \mathbb{R}^{+} \tag{3.74}$$

besagt, dass der zugrunde liegende Zufallsprozess, in diesem Fall der a-priori Schätzfehler, weiß ist. Analog zur Bedeutung eines weißen Prozesses am Ausgang

[5] Weist die Matrix $\mathbf{C}(k)$ nicht vollen Spaltenrang auf, kann die Pseudoinverse trotzdem berechnet werden. Dazu muss eine Eigenwertzerlegung $\mathbf{C}(k) = \mathbf{U}(k)\mathbf{D}(k)\mathbf{V}^{\mathrm{H}}(k)$ durchgeführt werden. Für die Pseudoinverse gilt dann: $\mathbf{C}^{+}(k) = \mathbf{V}(k)\mathbf{D}^{+}(k)\mathbf{U}^{\mathrm{H}}(k)$, wobei $\mathbf{D}^{+}(k)$ aus der Matrix $\mathbf{D}(k)$ durch Invertieren aller Elemente ungleich Null berechnet wird.

eines Prädiktorfehlerfilters, bedeutet dies, dass der auf Grundlage der zum Zeitpunkt $k-1$ vorliegenden Daten vorhergesagte a-priori Schätzwert $\hat{\mathbf{x}}(k|k-1)$ des Zustandsvektors $\mathbf{x}(k)$ diesen bestmöglich prädiziert hat. Alle vorliegenden Korrelationseigenschaften wurden demnach ausgenutzt und das Restfehlersignal ist somit statistisch unabhängig.

Des Weiteren erfüllt diese zwar die Bedingung aus Gleichung 3.67, nicht aber die zweite Symmetriebedingung aus Gleichung 3.70. Die Kalman-Verstärkung des hier verwendeten Kalman-Filters ist somit eine verallgemeinerte Inverse der Messmatrix $\mathbf{C}(k)$:

$$\mathbf{K}(k) \;=\; \mathbf{C}^{-}(k) = \mathbf{C}(k)\left(\mathbf{C}^{\mathrm{H}}(k)\mathbf{C}(k)\right)^{-1}. \tag{3.75}$$

Dieser Zusammenhang ist nicht allgemein gültig, sondern beruht auf der hier getroffenen Wahl des zugrunde liegenden Signalmodells und der Modellierung des Fahrzeuggeräuschs im Zustandsvektor anstatt durch weißes Messrauschen, was dem klassischen Ansatz entspräche.

3.4.2 Verbesserungen der Robustheit

Die Robustheit bezüglich der numerischen Stabilität und bezüglich der Entstehung von sogenannten im Ausgangssignal hörbaren Artefakten (sogenannten *Musical Tones*) des in den vorangegangenen Abschnitten hergeleiteten Kalman-Filter Algorithmus' lässt sich durch zwei Änderungen in dessen Ablauf verbessern. Diese werden im Folgenden beschrieben.

Nachglätten der geschätzten Sprachkomponente

Der geschätzte Zustandsvektor $\hat{\mathbf{x}}(k|k)$ enthält nicht nur den gesuchten Wert der Sprachkomponente zum Abtastzeitpunkt k, sondern auch die der vergangenen $p-1$ Abtastzeitpunkte. Dies sind:

$$\hat{s}(k),\ \hat{s}(k-1), \ldots,\ \hat{s}(k-p+1).$$

Ein beliebiger Schätzwert $\hat{s}(k_0)$ kann demnach erstmalig zum Zeitpunkt $k = k_0$ und letztmalig zum Zeitpunkt $k = k_0 + p - 1$ aus dem Zustandsvektor $\hat{\mathbf{x}}(k|k)$ extrahiert werden. Für die Herleitung des Kalman-Filters wurde, bezogen auf dieses Beispiel, stets der Abtastzeitpunkt $k = k_0$ verwendet.

Extrahiert man zu einem späteren Zeitpunkt als $k = k_0$, so wird der Schätzwert auf Grundlage der nachfolgenden Daten weiter verbessert, was hier als *Nachglätten* bezeichnet wird. Während in der Prädiktionsstufe des Kalman-Filters der gesuchte Wert $\hat{s}(k_0)$ nur um eine Position innerhalb des Zustandsvektors $\hat{\mathbf{x}}(k|k)$ nach

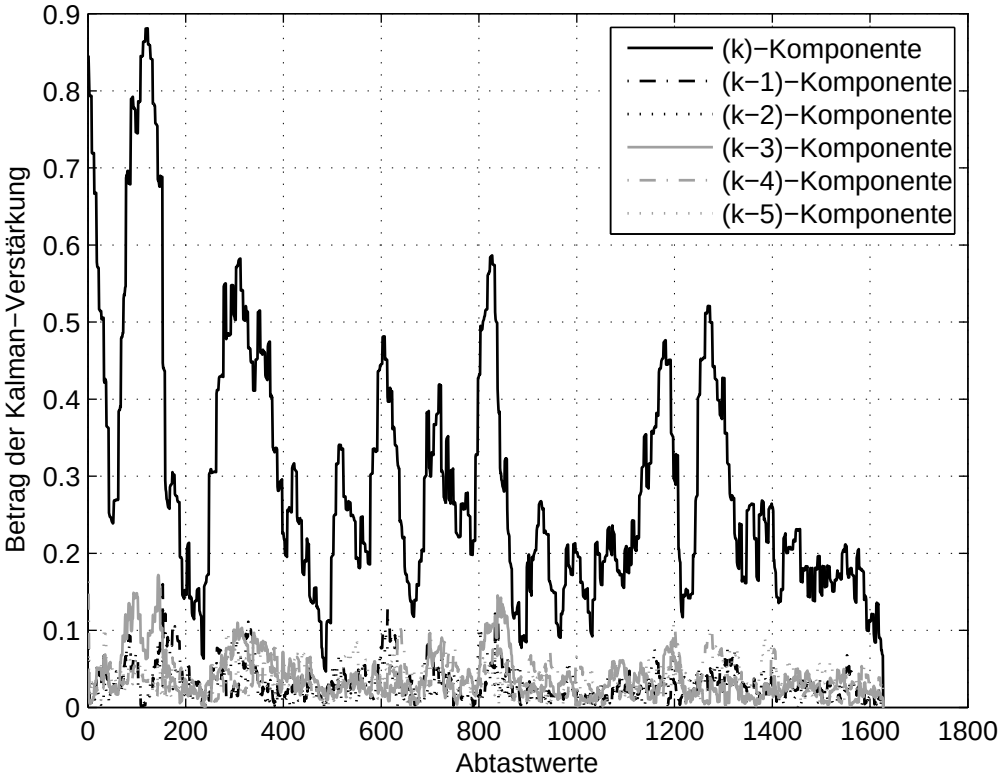

Abbildung 3.4: *Zeitlicher Verlauf des Betrags der einzelnen Komponenten der Kalman-Verstärkung für* $N = 1$ *und* $p = 6$*. Dabei sind nur diejenigen Komponenten abgebildet, die auf die Schätzwerte der Sprachquelle* $\hat{s}(k), \hat{s}(k-1), \ldots, \hat{s}(k-5)$ *wirken.*

oben geschoben wird (siehe Schiebematrix in Abschnitt 3.1.3), verursachen die nachfolgenden Messwerte $\mathbf{y}(k)$ für $k > k_0$ eine weitere Gewichtung des Schätzwertes durch die Kalmanverstärkung $\mathbf{K}(k)$. Diese ist zwar verglichen zu der beim Abtastzeitpunkt $k = k_0$ sehr gering, aber dennoch von Null verschieden, was in Abbildung 3.4 dargestellt ist. Daher wird durch Nachglätten eine Verbesserung des gesuchten Schätzwertes $\hat{s}(k_0)$ erreicht.

Diese Verbesserung geht allerdings zu Lasten der Laufzeit des Algorithmus', da mit jedem Nachglättungszyklus der gesuchte Schätzwert einen Abtastzeitpunkt später am Ausgang der Geräuschreduktion anliegt. Diese Problematik wird bei der Diskussion der Gesamtverzögerung in Abschnitt 5.2.3 behandelt.

Überschätzung des Geräuschs

Eine übliche Methode in Algorithmen zur Geräuschreduktion, um deren Robustheit bezüglich *Musical Tones* zu verbessern, besteht im sogenannten *Überschätzen* des Geräuschs. Darunter versteht man die künstliche Erhöhung der Geräuschleistung beispielsweise durch Multiplikation mit einem Faktor größer als Eins. Bei dem in dieser Arbeit vorgestellten Verfahren kann die Geräuschüberschätzung an drei Stellen angewendet werden, wobei die Motivation dazu jeweils unterschiedlich ist:

- Innerhalb des Kalman-Filters mittels des zusätzlich eingeführten Messrauschens $\mathbf{z}(k)$.
- Innerhalb des Kalman-Filters bei den geschätzten Geräuschleistungen der Anregungsprozesse $w_i(k)$ der Geräuschmodelle.
- Bei der Schätzung der AR-Koeffizienten innerhalb der DAKF-Methode, welche inklusive der dort notwendigen Geräuschüberschätzung in Abschnitt 4.3.1 beschrieben wird.

Der Rauschprozess $\mathbf{z}(k)$ aus der Messgleichung 3.25 taucht im Kalman-Filter nur in der Gleichung 3.48 für die Kalmanverstärkung in Form der Kovarianzmatrix $\mathbf{P}_{\mathbf{z}}(k)$ auf. Da $\mathbf{z}(k)$ aus N skalaren, weißen Rauschprozessen zusammengesetzt ist, weist $\mathbf{P}_{\mathbf{z}}(k)$ nur auf der Hauptdiagonalen Elemente ungleich Null auf:

$$\mathbf{P}_{\mathbf{z}}(k) \;=\; \begin{bmatrix} \sigma_{z_1}^2(k) & 0 & \cdots & 0 \\ 0 & \sigma_{z_2}^2(k) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_{z_N}^2 \end{bmatrix}. \tag{3.76}$$

Diese Varianzen erhöhen die Hauptdiagonale des zu invertierenden Terms aus Gleichung 3.48. Die Erhöhung der Hauptdiagonalen einer Matrix verbessert prinzipiell deren Konditionierung, was zu einer verbesserten numerischen Stabilität bei der Berechnung der Kalman-Verstärkung führt. Außerdem wird $\mathbf{K}(k)$ betragsmäßig kleiner, was bedeutet, dass dem aktuellen Messvektor $\mathbf{y}(k)$ weniger (und damit der Prädiktion mehr) vertraut wird. Dies bedeutet aber auch, dass der Algorithmus langsamer auf Änderungen reagiert, weshalb zwischen numerischer Stabilität und Konvergenzgeschwindigkeit abgewogen werden muss.

Da über $\mathbf{z}(k)$ keine Kenntnisse vorhanden sind, wird es zweckmäßigerweise konstant und zeitinvariant gewählt:

$$\sigma_{z_1}^2(k) = \sigma_{z_2}^2(k) = \cdots = \sigma_{z_N}^2(k) \;=\; \sigma_z^2. \tag{3.77}$$

Bei den dieser Arbeit zugrunde liegenden Audiodaten hat sich $\sigma_z^2 = 0,001/\gamma^2$ bewährt. Dabei bezeichnet γ einen Normierungsfaktor mit Betrag Eins und gleicher Dimension wie $z_i(k)$.

Das Überschätzen der Geräuschleistungen $\sigma^2_{w_i}(k)$ in der Matrix $\mathbf{P_u}(k)$ bewirkt hingegen eine Reduktion der im Kalman-Filter entstandenen Musical Tones [39]. Hierfür hat sich der für alle Geräuschquellen identische und über der Zeit konstante Faktor Zwei bewährt. In [39] wurde vorgeschlagenen, zwischen Sprachpausen und -phasen bei der Wahl dieses Faktors zu unterscheiden. Dies brachte im mehrkanaligen Fall keine hörbare Verbesserung, wobei der hier gefundene Wert zwischen den beiden dort vorgeschlagenen (3 in Sprachpausen und 1,5 in Sprachphasen) liegt.

3.5 Numerischer Aufwand

In diesem letzen Abschnitt des Kapitels wird der numerische Aufwand des hier verwendeten Kalman-Filter-Algorithmus' abgeschätzt. Dabei werden implementierungsspezifische Aspekte wie zum Beispiel die aus der Filterbankstruktur resultierenden komplexwertigen Signale berücksichtigt.

Wie bereits zuvor erwähnt, beschränkt sich die Darstellung in diesem Kapitel auf das Kalman-Filter selbst. Der numerische Aufwand für die Schätzverfahren zur Parametrierung des Filters wird im nächsten Kapitel behandelt.

Für die folgenden Überlegungen wird angenommen, dass eine M-kanalige Filterbank verwendet wird, wobei M gerade ist. Wie im Kapitel 5 noch näher erläutert wird, bedeutet dies im Fall der hier vorliegenden reellen Eingangssignale, dass im Teilbandbereich zwei reellwertige und $M/2-1$ komplexwertige Frequenzbänder vorliegen. Der Einfachheit halber werden für alle Bänder die gleichen konstanten Ordnungen für die Sprachquelle (Ordnung p) und die N Geräuschquellen (Ordnung q) benutzt.

Im Folgenden werden die vorkommenden Matrizen und Vektoren zunächst so behandelt, als ob diese vollständig besetzt wären. Das heißt, es wird für alle Elemente angenommen, dass sie ungleich Null sind. Aufgrund dieser Annahme hängt die Anzahl der Multiplikationen nur von den Dimensionen der verwendeten Größen ab, welche in Tabelle 3.2 aufgelistet sind.

Die Anzahl der notwendigen Multiplikation für die Berechnung des Produkts zweier beliebig dimensionierter vektorieller Größen Θ und Ψ ergibt sich zu:

$$\text{Anz. Multipl.} = (\text{Zeilenanz. } \Theta) \cdot (\text{Zeilenanz. } \Psi) \cdot (\text{Spaltenanz. } \Psi). \qquad (3.78)$$

Mit Hilfe von Gleichung 3.78 sowie den Tabellen 3.1 und 3.2 kann nun der numerische Aufwand für die notwendigen Multiplikationen des Kalman-Filter-Algorithmus' bestimmt werden. Dieser ist für die einzelnen Berechnungsschritte der Prädiktions- und Korrektur-Stufe in Tabelle 3.3 aufgeführt. Die Initialisier-

Tabelle 3.2: *Dimension der vorkommenden vektoriellen Größen.*

Bezeichner	Dimension (Zeilen×Spalten)
$\hat{\mathbf{x}}(k\|k-1)$, $\hat{\mathbf{x}}(k\|k)$	$(p+Nq)\times 1$
$\mathbf{A}(k\|k-1)$	$(p+Nq)\times(p+Nq)$
$\mathbf{P_e}(k\|k-1)$, $\mathbf{P_e}(k\|k)$	$(p+Nq)\times(p+Nq)$
$\mathbf{B}$	$(p+Nq)\times(N+1)$
$\mathbf{P_u}(k)$	$(N+1)\times(N+1)$
$\mathbf{K}$	$(p+Nq)\times N$
$\mathbf{C}(k)$	$(p+Nq)\times N$
$\mathbf{P_z}(k)$	$N\times N$
$\mathbf{y}(k)$	$N\times 1$

ungs-Stufe ist für den Gesamtaufwand vernachlässigbar, da sie nur einmal durchlaufen wird. Aufgrund der zuvor getroffenen Annahme der vollbesetzten Matrizen und Vektoren stellen die Tabellenwerte den *Worst Case* dar, der die besonderen Eigenschaften der verwendeten Matrizen unberücksichtigt lässt. Die zu invertierende Matrix bei der Berechnung der Kalman-Verstärkung weist im Allgemeinen keine spezielle Struktur auf, so dass der Aufwand zur Invertierung mit N^3 abgeschätzt wurde.

Tabelle 3.3: *Abschätzung der maximal zu erwartenden Anzahl von Multiplikationen für jeden Berechnungsschritt pro Frequenzband.*

Gleichung	Anzahl Multiplikationen
A-priori Schätzwert	$(p+Nq)^2$
A-priori Kovarianz	$2(p+Nq)^3+(p+Nq)(N+1)^2+(N+1)(p+Nq)^2$
Kalmanverstärkung	$2N(p+Nq)^2+2(p+Nq)N^2+N^3$
A-posteriori Schätzwert	$2N(p+Nq)$
A-posteriori Kovarianz	$(p+Nq)^3+N(p+Nq)^2$

Diese Multiplikationsanzahlen gelten für jedes Frequenzband. Da reelle Eingangssignale bei der Filterbank angenommen wurden, müssen aufgrund der Symmetrie nur $M/2+1$ dieser Bänder berücksichtigt werden, wobei $M/2-1$ von ihnen komplexwertige Signale enthalten. Da eine komplexwertige Multiplikation im Allgemeinen vier[6] reellen entspricht, müssen die Tabellenwerte in diesen Bändern entsprechend um den Faktor vier zusätzlich vergrößert werden.

Berücksichtigt man darüber hinaus die besondere Struktur der Matrizen, so reduziert sich der Aufwand und man erhält eine verbesserte Abschätzung der notwen-

[6]Neben der allgemeinen Form $z = z_1 \cdot z_2 = (a+jb)(c+jd) = (ac-bd)+j(ad+bc)$, die vier Multiplikationen und drei Additionen benötigt, gibt es eine 2-Schritt-Lösung, die nur drei reelle Multiplikationen benötigt: 1.) $\Re\{z\} = ac-bd$; 2.) $\Im\{z\} = (a+b)(c+d)-\Re\{z\}$.

digen Multiplikationen. Die Matrizen $\mathbf{A}(k|k-1)$, $\mathbf{C}(k)$, $\mathbf{P_u}(k)$ und $\mathbf{B}$ enthalten Elemente, die zu jedem Zeitpunkt k identisch Null sind. Dies betrifft insbesondere die $\mathbf{A}$-Matrix, deren Einträge zu mehr als der Hälfte Null sind (siehe Gleichungen 3.10 und 3.19). Werden zusätzlich alle Multiplikationen mit Eins vernachlässigt, weil diese in Compilern nicht als echte Multiplikation ausgeführt werden, so erhält man den in Tabelle 3.4 aufgeführten Aufwand. Dabei wurde davon ausgegangen, dass $\mathbf{C}(k)$ vollständig, das heißt mit allen möglichen Kreuzfilterfunktionen, besetzt ist.

Tabelle 3.4: *Abschätzung der maximal zu erwartenden Anzahl von Multiplikationen für jeden Berechnungsschritt pro Frequenzband unter Berücksichtigung der besonderen Struktur der verwendeten Matrizen.*

Gleichung	**Anzahl Multiplikationen**
A-priori Schätzwert	$(p+Nq)$
A-priori Kovarianz	$2(p+Nq)^2$
Kalmanverstärkung	$2N(p+Nq)(p+(N-1)q)+(p+Nq)N^2\cdots$ $+(p+(N-1)q)N^2+N^3$
A-posteriori Schätzwert	$N(p+Nq)+N(p+(N-1)q)$
A-posteriori Kovarianz	$(p+Nq)^3+(p+Nq)N(p+(N-1)q)$

Die Anzahl der Multiplikationen für ein reelles Teilband ist für jeweils beide Abschätzungen und verschiedene Mikrofonanzahlen sowie verschiedenen Ordnungen des Sprachmodells in Abbildung 3.5 dargestellt. Bei der tatsächlichen Implementierung können für jedes Frequenzband verschiedene Ordnungen für Sprache (p) und Geräusch (q) verwendet werden.

Zusammengefasst kann gesagt werden, dass der numerische Aufwand bezüglich der Multiplikationen jeweils kubisch mit der Mikrofonanzahl sowie den Modellordnungen wächst. Dieser Zusammenhang muss bei der Wahl der Parameter N, p und q berücksichtigt werden und gegen die erzielbare Verbesserung, die durch eine Vergrößerung dieser Parameter bewirkt werden kann, abgewogen werden.

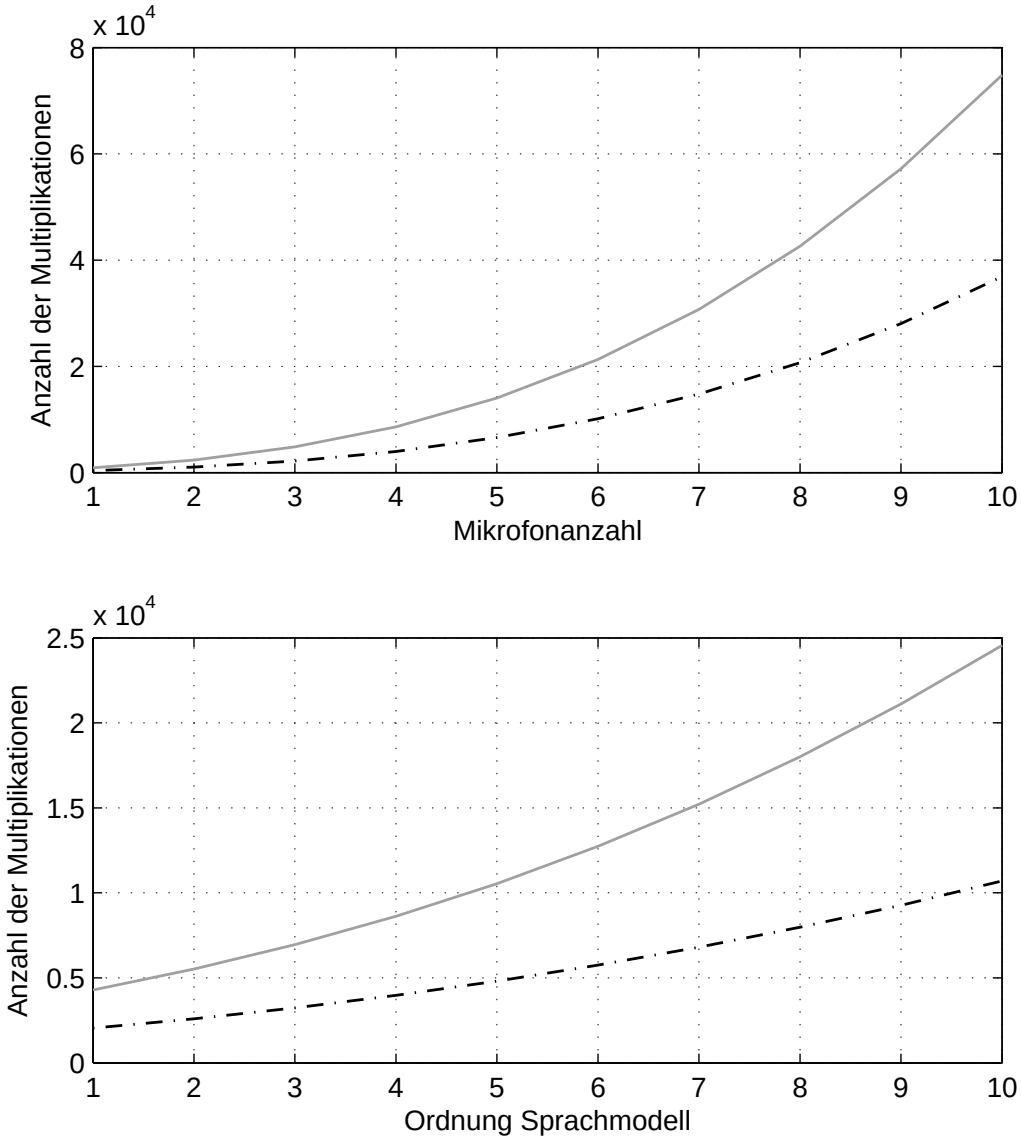

Abbildung 3.5: *Numerischer Aufwand eines reellwertigen Frequenzbands für verschiedene Mikrofonanzahlen N bei konstanter Ordnung $p = 4$ (oben), sowie für verschiedene Ordnungen p bei $N = 4$ Mikrofonen (unten). Dargestellt sind der Aufwand für $q = 2$ bei Worst-Case-Abschätzung (grau), sowie unter Berücksichtigung der besonderen Struktur der Matrizen (schwarz gestrichelt).*

Kapitel 4

Schätzung der Filterparameter

Im vorangegangenen Kapitel 3 wurde ein auf Kalman-Filterung basiertes Verfahren vorgestellt. Um es verwenden zu können, müssen alle darin vorkommenden Parameter geschätzt werden. Im Einzelnen sind dies:

- die AR-Koeffizienten des Sprachsignals $a_{s,i}(k)$,
- die AR-Koeffizienten der Geräuschsignale $a_{n_1,i}(k), a_{n_2,i}(k), \ldots, a_{n_N,i}(k)$
- die Anregungsleistung des Sprachmodells $\sigma_s^2(k)$,
- die Anregungsleistungen des Geräuschmodells $\sigma_{n_1}^2(k), \sigma_{n_2}^2(k), \ldots, \sigma_{n_N}^2(k)$,
- die Raumimpulsantworten $h_{1,i}(k), h_{2,i}(k), \ldots, h_{N,i}(k)$ sowie
- die für das jeweilige Modell gewählten (siehe Abschnitt 3.3) Filterfunktionen $g_{t_1 t_2,i}(k)$ mit $t_1, t_2 \in \{1, \ldots, N\}$ und $t_1 \neq t_2$.

Die zu bestimmenden Parameter können unterteilt werden in diejenigen der Systemgleichung 3.23, welche die Entstehung der Sprach- und Geräuschsignale des Kalman-Filters beschreiben, und diejenigen der Messgleichung 3.24, welche den Raum, der die Mikrofone umgibt, charakterisieren. Allen gemein ist, dass sie zeitvariant sind und somit in regelmäßigen Abständen, das heißt alle 20 bis 50 ms (siehe Abschnitt 2.1.1), erneut bestimmt werden müssen.

In diesem Kapitel werden nun Verfahren vorgestellt, die es ermöglichen, diese zeitvarianten Parameter zu schätzen. Hauptaugenmerk liegt hierbei auf der Schätzung der AR-Parameter, da sie für die Leistungsfähigkeit des Kalman-Filters von entscheidender Bedeutung sind. Unter dem Begriff *AR-Parameter* werden in diesem Zusammenhang sowohl die AR-Koeffizienten, beispielsweise $a_{s,i}(k)$, als auch die zugehörige Anregungsleistung, in diesem Beispiel $\sigma_v^2(k)$, zusammengefasst.

Begonnen wird diese Kapitel mit einer kurzen mathematischen Beschreibung des AR-Modells und der Lösung des sogenannten *Anpassungsproblems*. Dabei wird insbesondere auf den engen Zusammenhang zwischen AR- und Prädiktor-

koeffizienten eingegangen, da letztere in Kapitel 3 zur Repräsentation des AR-Modells im Kalman-Filter verwendet wurden. Im Hinblick auf die hier vorgeschlagenen Verfahren werden sowohl Methoden der parametrischen als auch der nicht-parametrischen Spektralschätzung benötigt. Diese werden im zweiten Teil des Kapitels vorgestellt und bezüglich der Verwendbarkeit zur Schätzung der AR-Parameter bewertet. Daran anschließend werden im dritten Teil verschiedene für den mehrkanaligen Fall erweiterte Verfahren vorgestellt. Ausgegangen wird dabei jeweils von der in [38, 39] vorgeschlagenen einkanaligen Struktur, welche ebenfalls kurz erläutert und zudem in Kapitel 6 als Referenz dienen wird. Der letzte Teil dieses Kapitels befasst sich mit der Schätzung der Raumimpulsantworten und Filterfunktionen.

Wie im vorangegangenen Kapitel 3 werden alle Herleitungen im Vollband unter expliziter Annahme von komplexwertigen Größen durchgeführt, so dass die gefundenen Gleichungen später direkt im Teilband verwendet werden können.

4.1 Grundlagen der AR-Modellierung

In Kapitel 3 wurden die Eingangssignale des Kalman-Filters als Zufallsprozesse modelliert. Zur deren Beschreibung werden daher über die Schar der Realisierungen berechnete Momente, die sogenannten *Scharmittelwerte*, wie zum Beispiel die Autokorrelationsfunktion verwendet. In der Wirklichkeit müssen diese durch Zeitmittelwerte ersetzt werden, wozu Ergodizität modelliert werden muss, welche wiederum die Stationarität der betroffenen Zufallsprozesse bedingt. Analog zum Begriff der schwachen Stationarität (siehe Abschnitt 2.1.1) wird von schwacher Ergodizität gesprochen, falls die betroffenen Zufallsprozesse nur für Momente bis zur Ordnung zwei ergodisch sind. Wie bereits in Abschnitt 2.1.1 beschrieben, ist die allgemeine Annahme von Stationarität für Sprachsignale unrealistisch. Daher werden im Folgenden sogenannte *Kurzzeit-Verfahren* verwendet, um in den als stationär angenommenen Sprachsegmenten die AR-Parameter zu bestimmen.

4.1.1 Mathematische Beschreibung des AR-Modells

Das autoregressive oder einfach AR-Modell ist ein Sonderfall der spektralen Modellierung eines Signals durch eine gebrochen rationale Funktion im z-Bereich. Ausgehend von der allgemeinen Differenzengleichung

$$y(k) + \sum_{i=1}^{m} a_i y(k-i) \;=\; \sum_{i'=0}^{m'} b_{i'} v(k-i') \tag{4.1}$$

im Zeitbereich, erhält man durch Transformation in den z-Bereich und Ergänzung der linken Summe aus Gleichung 4.1 um den Vorfaktor von $y(k)$, also $a_0 = 1$:

$$Y(z) = \underbrace{\frac{\sum_{i'=0}^{m'} b_{i'} z^{-i'}}{\sum_{i=0}^{m} a_i z^{-i}}}_{H_{\mathrm{ARMA}}(z)} V(z). \tag{4.2}$$

Hierbei bezeichnen $v(k)$ bzw. $V(z)$ den als weiß angenommenen Eingangsprozess mit der Varianz σ_v^2 und $y(k)$ bzw. $Y(z)$ den entsprechenden Ausgangsprozess. Dies ist die allgemeine Form eines digitalen Filters. Für $b_{i'} = 0$ unterscheidet sich Gleichung 4.1 von den Gleichungen 3.11 bis 3.13 im Vorzeichen der linken Summe. Auf diesen Unterschied wird im Verlauf des Abschnitts 4.1.1 noch eingegangen. Der Zusammenhang zwischen Ein- und Ausgang wird durch die aus einem Zählerpolynom $B(z)$ und Nennerpolynom $A(z)$ bestehende Übertragungsfunktion

$$H_{\mathrm{ARMA}}(z) = \frac{B(z)}{A(z)} = \frac{\sum_{i'=0}^{m'} b_{i'} z^{-i'}}{\sum_{i=0}^{m} a_i z^{-i}} = \frac{\prod_{i'=0}^{m'} (z - z_{0,i'})}{\prod_{i=0}^{m} (z - z_{\mathrm{P},i})} \tag{4.3}$$

beschrieben, welche vollständig durch die Nullstellen $z_{0,i'}$ sowie die Polstellen $z_{\mathrm{P},i}$, das sind die Nullstellen von $A(z)$, definiert ist. Hierbei bezeichnen m und m' die Ordnungen des rekursiven bzw. transversalen Anteils.

Dieses Modell wird als autoregressives moving-average (ARMA) oder Pol-Nullstellen-Modell bezeichnet und weist zwei Sonderfälle auf:

- Für $a_i = 0$ mit $i = 1, \ldots, m$ verschwindet das Nennerpolynom, da $a_0 = 1$ angenommen wurde. Man erhält eine rein transversale Übertragungsfunktion, die ausschließlich Nullstellen enthält. Daher wird dieser Fall als Nullstellen- oder moving-average (MA) Modell bezeichnet. Es gilt:
  $H_{\mathrm{MA}}(z) = \sum_{i'=0}^{m'} b_{i'} z^{-i'}$.
- Für $b_{i'} = 0$ mit $i' = 1, \ldots, m'$ verschwindet das Zählerpolynom und die Übertragungsfunktion wird rein rekursiv. In diesem Fall wird das Modell ausschließlich über die Polstellen definiert und man spricht von einem Polstellen- oder autoregressiven (AR) Modell. Es gilt:
  $H_{\mathrm{AR}}(z) = 1 / \sum_{i=0}^{m'} a_i z^{-i'}$.

Die Grundidee bei der Verwendung dieser Modelle zur Spektralschätzung ist, anstatt das Spektrum eines Signals direkt zu schätzen, das gegebene Modell auf dieses Signal anzupassen. Dadurch müssen hier im Vergleich zu den direkten Methoden nur die wenigen Parameter des jeweiligen Modells geschätzt werden. Dabei gilt grundsätzlich, dass die Ordnungen m und m' umso größer sein müssen, desto komplexer die abzubildende spektrale Struktur ist.

In dem Gebiet der Sprachsignalverarbeitung spielt das AR-Modell die bedeutendste Rolle der drei vorgestellten Modelle, weshalb im weiteren Verlauf aus-

schließlich dieses untersucht wird. Seine Bedeutung gründet sich auf einigen Vorteilen, die es gegenüber den anderen Modellen aufweist.

Zum einen ergeben die Polstellen $z_{\mathrm{P},i}$ der Übertragungsfunktion im Spektrum schmale, scharf ausgeprägte Maxima, während die spektralen Täler nicht direkt modelliert werden und somit weniger scharf ausfallen. Beim MA-Modell ist es genau umgekehrt. Dort modellieren die Nullstellen $z_{0,i}$ scharf begrenzte spektrale Täler, wodurch die Maxima relativ breit ausfallen. Aus diesem Grund eignet sich das AR-Modell sehr gut als Formfilter zur Modellierung der Pitchstruktur von Sprachsignalen in dem in Abschnitt 2.3 vorgestellten Quelle-Filter-Modell.

Der Hauptvorteil liegt hier aber darin, dass es eng mit dem linearen Prädiktor verwandt ist. Die Anpassung eines AR-Modells auf ein gegebenes Signal führt zu dem gleichen linearen Optimierungsproblem wie bei der linearen Prädiktion. Dies schließt auch die darin vorkommenden Größen wie zum Beispiel die Autokorrelationsfunktion ein. Die Lösung basiert auf der Minimierung des mittleren quadratischen Fehlers, wofür effiziente Verfahren wie zum Beispiel der Levinson-Durbin-Algorithmus zur Verfügung stehen. Die Anpassung der anderen beiden Modelle ist im Allgemeinen deutlich aufwendiger, da nichtlineare Gleichungssysteme gelöst werden müssen [26, 53], weshalb die fehlende Möglichkeit der Nasaltraktmodellierung (siehe Abschnitt 2.3) beim AR-Modell in Kauf genommen wird.

Stabilitätsbetrachtungen bei AR-Modellen

Ein Nachteil, der sich besonders gegenüber dem transversalen MA-Modell auswirkt, ist die Tatsache, dass rekursive Modelle prinzipiell instabil werden können. Es werden daher Kriterien benötigt, die es ermöglichen, die Stabilität zu überprüfen bzw. sicherzustellen.

Folgende Stabilitätsbedingungen, von denen einige erst im Verlauf dieses Kapitels näher beschrieben werden, sind äquivalent:

- Alle Polstellen liegen innerhalb des Einheitskreises. Es gilt also: $|z_{\mathrm{P},i}| < 1$ für alle $i = 1, \ldots, m$. In diesem Fall konvergiert $H_{\mathrm{AR}}(z)$ auf dem Einheitskreis[1] und es existiert die Fourier-Transformierte $H_{\mathrm{AR}}(\Omega) = H_{\mathrm{AR}}(z)|_{z=e^{j\Omega}}$ des Formfilters auf dem Einheitskreis.

[1] Aus der Lage der Polstellen in der komplexen Ebene kann man den ungefähren Verlauf des Frequenzgangs ablesen. Dazu geht man den Einheitskreis beginnend von der positiven realen Achse in mathematisch positiver Richtung ab. Eine komplette Umrundung entspricht dabei dem Durchlauf der Frequenz von $\Omega = 0$ bis $\Omega = 2\pi$. Beim Abgehen ergeben nun Pole, die dicht beim Einheitskreis liegen große Frequenzgangsmaxima, während weiter entfernte Pole nur kleinere Maxima erzeugen. Die Frequenz dieser Maxima bestimmt man anhand der Position auf dem Einheitskreis.

- Die Autokorrelationsmatrix $\mathbf{R}_{yy}$ ist positiv definit, also invertierbar.
- Das Leistungsdichtespektrum ist für alle Frequenzen stets nicht negativ. Das heißt, es gilt: $S_{yy}(\Omega) \geq 0$ für $0 \leq \Omega \leq 2\pi$.
- Der Betrag der Reflexionskoeffizienten (siehe Abschnitt 4.2.3) ist für alle zugelassenen i stets kleiner als Eins. Das heißt: $|\Gamma_i| < 1$.

4.1.2 Lösung des Anpassungsproblems

Zur Anpassung des AR-Modells an den Zufallsprozess $y(k)$ bei Anregung durch einen weißen Zufallsprozess $v(k)$ wird in Gleichung 4.1 $m'=0$ gesetzt. Zusätzlich wird der Faktor $b_0 = 1$ angenommen. Man erhält die Differenzengleichung des AR-Modells:

$$y(k) \quad = \quad v(k) - \sum_{i=1}^{m} a_i y(k-i). \tag{4.4}$$

Diese unterscheidet sich im Vorzeichen der Summe von dem in Gleichung 3.1 formulierten Ansatz, bei welchem Prädiktorkoeffizienten benutzt wurden. Sowohl der Prädiktionsansatz aus Gleichung 3.1 als auch der klassische AR-Modell-Ansatz aus Gleichung 4.4 führen zu Modellen mit identischen Eigenschaften. Der einzige Unterschied besteht darin, dass die Koeffizienten dieser Modelle jeweils zueinander negierte Vorzeichen haben. Daher wird an dieser Stelle zunächst der AR-Modell-Ansatz weitergeführt und später der Unterschied zum Prädiktionsansatz aufgezeigt.

Um die optimalen AR-Koeffizienten a_i zu berechnen, wird die Autokorrelationsfunktion des Ausgangs bestimmt und Gleichung 4.4 einmal eingesetzt. Dabei wird angenommen, dass alle vorkommenden Zufallsprozesse stationär sind. Diese Annahme dient lediglich der Vereinfachung der Herleitung und wird im nächsten Abschnitt wieder fallengelassen:

$$\begin{aligned}
r_{yy}(l) &= \mathrm{E}\left\{y^*(k)y(k+l)\right\} \\
&= \mathrm{E}\left\{y^*(k)\left(v(k+l) - \sum_{i=1}^{m} a_i y(k+l-i)\right)\right\} \\
&= \mathrm{E}\left\{y^*(k)v(k+l)\right\} - \sum_{i=1}^{m} a_i \mathrm{E}\left\{y^*(k)y(k+l-i)\right\} \\
&= r_{yv}(l) - \sum_{i=1}^{m} a_i r_{yy}(l-i).
\end{aligned} \tag{4.5}$$

Die Kreuzkorrelationsfunktion $r_{yv}(l)$ wird nun als Faltung des weißen Anregungsprozesses $v(k)$ und der zu bestimmenden Impulsantwort des als kausal vorausge-

setzten AR-Modells ausgedrückt:

$$r_{vy}(l) = \sigma_v^2 h(l) \quad \text{bzw.} \quad r_{yv}^*(-l) = \sigma_v^2 h^*(-l). \tag{4.6}$$

Aufgrund der angenommenen Kausalität und der rein rekursiven Struktur des AR-Modells in Gleichung 4.1 erhält man schließlich [46]:

$$r_{yy}(l) + \sum_{i=1}^{m} a_i r_{yy}^*(l-i) \;=\; \begin{cases} \sigma_v^2 & l = 0 \\ 0 & l > 0 \end{cases} . \tag{4.7}$$

Gleichung 4.7 kann als lineares Gleichungssystem geschrieben werden, und es ergibt sich das sogenannte *Yule-Walker*- oder auch *Normalen*-Gleichungssystem, dessen Zeilen jeweils Gleichung 4.7 für einen Wert von l (in aufsteigender Reihenfolge) entsprechen:

$$\left[\begin{array}{cccc} r_{yy}(0) & r_{yy}(-1) & \cdots & r_{yy}(-m) \\ \hdashline r_{yy}(1) & r_{yy}(0) & \cdots & r_{yy}(-m+1) \\ \vdots & \vdots & \ddots & \vdots \\ r_{yy}(m) & r_{yy}(m-1) & \cdots & r_{yy}(0) \end{array}\right] \left[\begin{array}{c} 1 \\ \hdashline a_1 \\ \vdots \\ a_m \end{array}\right] = \left[\begin{array}{c} \sigma_v^2 \\ \hdashline 0 \\ \vdots \\ 0 \end{array}\right]. \tag{4.8}$$

Dieses kann, wie durch die gestrichelte Linie angedeutet, in zwei Teile aufgespalten werden. Aus dem unteren Teil, also für $l = 1, \ldots, m$, erhält man nach Subtraktion der linken Spalte das lineare Gleichungssystem:

$$\left[\begin{array}{cccc} r_{yy}(0) & r_{yy}(-1) & \cdots & r_{yy}(-m+1) \\ r_{yy}(1) & r_{yy}(0) & \cdots & r_{yy}(-m+2) \\ \vdots & \vdots & \ddots & \vdots \\ r_{yy}(m-1) & r_{yy}(m-2) & \cdots & r_{yy}(0) \end{array}\right] \left[\begin{array}{c} a_1 \\ a_2 \\ \vdots \\ a_m \end{array}\right] = \left[\begin{array}{c} -r_{yy}(1) \\ -r_{yy}(2) \\ \vdots \\ -r_{yy}(m) \end{array}\right], \tag{4.9}$$

aus welchem die AR-Koeffizienten $a_1, ..., a_m$ bestimmt werden können, sofern die Werte von $r_{yy}(l)$ für $0 < l \leq m$ bekannt sind.

Gleichung 4.9 kann folgendermaßen in Vektor/Matrix-Schreibweise formuliert werden:

$$\mathbf{R}_{yy}\mathbf{a} \;=\; -\tilde{\mathbf{r}}_{yy}, \tag{4.10}$$

wobei $\mathbf{R}_{yy}$ die Autokorrelationsmatrix, $\mathbf{a}$ den Vektor der AR-Parameter und $\tilde{\mathbf{r}}_{yy}$ den Vektor der um eins verschobenen Elemente der Autokorrelationsfunktion bezeichnet[2]. Unter der Voraussetzung, dass die Autokorrelationsmatrix $\mathbf{R}_{yy}$ invertierbar ist, kann Gleichung 4.10 gelöst werden:

$$\mathbf{a} \;=\; -\mathbf{R}_{yy}^{-1}\tilde{\mathbf{r}}_{yy}. \tag{4.11}$$

[2]Man beachte, dass durch die anfangs angenommene Stationarität alle Größen noch ohne Zeitindex geschrieben werden.

Durch diese als *Yule-Walker-Gleichung* bezeichnete Lösung wird das Problem der Anpassung des AR-Modells auf das Signal $y(k)$ reduziert auf die Schätzung der Autokorrelationsfunktion $r_{yy}(l)$.

Die zu invertierende Autokorrelationsmatrix $\mathbf{R}_{yy}$ aus Gleichung 4.9 weist eine konjugiert-symmetrische Toeplitz-Struktur auf. Darunter versteht man eine Matrix, deren Elemente innerhalb einer Diagonalen konstant sind. Darüber hinaus ist die Autokorrelationsmatrix stets positiv semidefinit [19, 26, 46]. Für praktische Anwendungen kann sie als invertierbar angenommen werden, da Nicht-Invertierbarkeit nur in wenig realistischen Situation, wie zum Beispiel bei Anregung mit einem konstanten Signal, erreicht wird. Zur Inversion stehen effiziente Verfahren wie z.B. der Levinson-Durbin-Algorithmus (siehe Abschnitt 4.2.3) zur Verfügung, deren Aufwand nicht kubisch mit m^3, sondern nur quadratisch mit m^2 wächst [19].

Aus der ersten Zeile des Yule-Walker-Gleichungssystems aus Gleichung 4.8 kann mit Hilfe der Lösung aus Gleichung 4.11 schließlich die noch fehlende Anregungsleistung σ_v^2 bestimmt werden:

$$\sigma_v^2 \;=\; r_{yy}(0) + \tilde{\mathbf{r}}_{yy}^{\mathrm{H}}\mathbf{a} = r_{yy}(0) - \tilde{\mathbf{r}}_{yy}^{\mathrm{H}}\mathbf{R}_{yy}^{-1}\tilde{\mathbf{r}}_{yy}. \tag{4.12}$$

4.1.3 Zusammenhang mit linearer Prädiktion

Der Zusammenhang mit dem linearen Prädiktor soll nun kurz erläutert werden. Ein linearer Prädiktor (LP) ist eine Rechenvorschrift, die auf Grundlage von vergangenen Signalwerten zukünftige schätzt [18, 19, 26, 52, 53, 21, 46]:

$$\hat{y}(k) \;=\; \sum_{i=1}^{m} a_{\mathrm{LP},i}\, y(k-i). \tag{4.13}$$

Genau genommen kann ein linearer Prädiktor auch Schätzwerte für vergangene Signalwerte, die bereits nicht mehr im Speicher sind, berechnen. Daher spricht man in diesem Zusammenhang von Vorwärts- und Rückwärtsprädiktion [19].

Die Güte der Schätzung aus Gleichung 4.13 ist unter anderem von der gewählten Ordnung, hier ebenfalls mit m notiert, abhängig. Die Struktur des linearen Prädiktors und des Prädiktorfehlerfilters sind in Abbildung 4.1 dargestellt. Dabei bezeichnet $e^{(\mathrm{f},m)}(k)$ den Vorwärts-Prädiktionsfehler der Ordnung m.

Für die Übertragungsfunktion eines linearen Prädiktors der Ordnung m und den Koeffizienten $a_{\mathrm{LP},i}$ gilt:

$$H_{\mathrm{LP}}(z) \;=\; \sum_{i=1}^{m} a_{\mathrm{LP},i} z^{-i}. \tag{4.14}$$

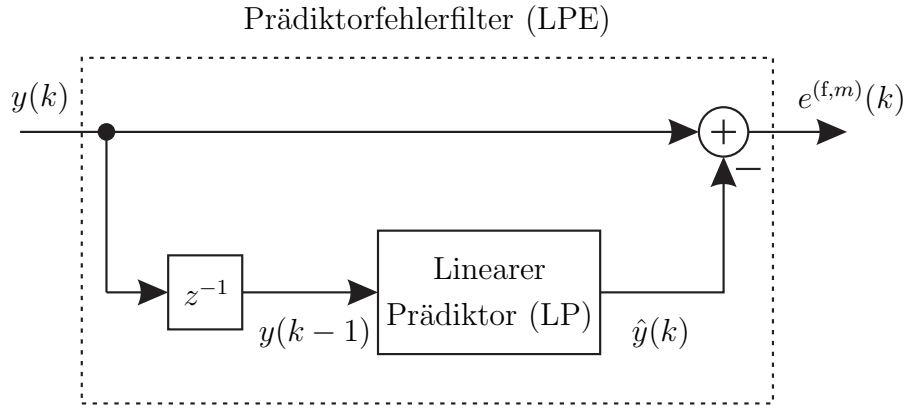

Abbildung 4.1: *Struktur des linearen Prädiktors (LP) und des Prädiktorfehlerfilters (LPE).*

Betrachtet man nicht nur das prädizierte Signal, sondern die Differenz zwischen wahrem und vorhergesagtem Signalwert, dies entspricht dem Vorwärts-Prädiktionsfehler, ergibt sich das sogenannte Prädiktorfehlerfilter (LPE), welches durch den gestrichelten Kasten in Abbildung 4.1 dargestellt ist. Dessen Koeffizienten erhält man durch die Negation der in Gleichung 4.14 gefunden Prädiktorkoeffizienten:

$$H_{\mathrm{LPE}}(z) \;=\; 1 - \sum_{i=1}^{m} a_{\mathrm{LP},i} z^{-i} = 1 + \sum_{i=1}^{m} \left(-a_{\mathrm{LP},i}\right) z^{-i}. \tag{4.15}$$

Die in Gleichung 4.15 enthaltene Summe kann auch zu einem Satz Koeffizienten zusammengefasst werden:

$$H_{\mathrm{LPE}}(z) \;=\; \sum_{i=0}^{m} a_{\mathrm{LPE},i} z^{-i}, \tag{4.16}$$

mit:

$$a_{\mathrm{LPE},i} \;=\; \begin{cases} 1 & \text{für } i = 0 \\ -a_{\mathrm{LP},i} & \text{für } 0 < i \leq m \end{cases}. \tag{4.17}$$

Der in Gleichung 4.17 dargestellte Koeffizientenvektor:

$$\mathbf{a}_{\mathrm{LPE}} \;=\; \left[1, a_{\mathrm{LPE},1}, a_{\mathrm{LPE},2}, \ldots, a_{\mathrm{LPE},m}\right]^{\mathrm{T}} \tag{4.18}$$

entspricht dabei dem Vektor, mit dem die Matrix im Yule-Walker-Gleichungssystem in Gleichung 4.8 multipliziert wird. Das AR-Modell ist im z-Bereich gegeben durch die Transformation der Differenzengleichung 4.1 mit $b_i = 0$ für $i > 0$:

$$H_{\mathrm{AR}}(z) \;=\; \frac{B(z)}{A(z)} = \frac{b_0}{1 + \sum_{i=1}^{m} a_i z^{-i}}. \tag{4.19}$$

Auch hier liefert ein Vergleich mit $H_{\mathrm{LPE}}(z)$ aus Gleichung 4.15 für $b_0 = 1$ sofort:

$$H_{\mathrm{AR}}(z) \;=\; \frac{1}{H_{\mathrm{LPE}}(z)}. \tag{4.20}$$

Das Prädiktorfehlerfilter ist demnach bis auf einen reellen Verstärkungsfaktor b_0 das inverse Filter zu dem gesuchten AR-Formfilter, und es weist die gleichen Koeffizienten auf. In diesem Zusammenhang spricht man auch von Analysefilter (LPE) und Synthesefilter (AR). Da $H_{\mathrm{LPE}}(z)$ stets minimalphasig ist [26], folgt daraus die Stabilität des Synthesefilters $H_{\mathrm{AR}}(z)$. Daher können alle Schätzmethoden für die lineare Prädiktion auch für die Schätzung von AR-Koeffizienten verwendet werden. Der Zusammenhang zwischen den Koeffizienten des linearen Prädiktors und denen des AR-Modells ergibt sich aus den Gleichungen 4.17 und 4.20 schließlich zu [53]:

$$a_i = a_{\mathrm{LPE},i} = -a_{\mathrm{LP},i} \quad \text{für} \quad i = 1, \ldots, m. \tag{4.21}$$

Aufgrund des in den Gleichungen 3.1 bis 3.3 gewählten Ansatzes, muss für die AR-Koeffizienten $a_{s,i}(k)$ und $a_{n_l}(k)$ für $k = 1, \ldots, N$ des Kalman-Filters aus Kapitel 3 die Darstellung der Prädiktorkoeffizienten $a_{\mathrm{LP},i}(k)$ verwendet werden.

4.2 Kurzzeit-Spektralanalyse

Im vorangegangenen Abschnitt wurden die mathematischen Eigenschaften und Zusammenhänge des AR-Modells aufgezeigt. Dieser Teil des Kapitels befasst sich mit den grundlegenden parametrischen und nicht-parametrischen Methoden zur Spektralschätzung, soweit sie im hier vorgeschlagenen Verfahren Verwendung finden. Die zuvor getroffene Annahme der Stationarität der betroffenen Zufallsprozesse wird ab sofort fallen gelassen. Daher richtet sich das Hauptaugenmerk auf die Auswirkungen dieser Kurzzeit-Verfahren auf das Verhalten der geschätzten AR-Parameter.

Die Yule-Walker-Gleichung 4.11 reduziert die Schätzung der AR-Parameter auf die Schätzung der Autokorrelationsfunktion des Zielsignals. Aus diesem Grund werden zunächst Verfahren zur Kurzzeitschätzung der Autokorrelationsfolge eines Zufallsprozesses diskutiert.

Des Weiteren besteht aufgrund der Wiener-Khintchine-Beziehung über die Fourier-Transformation eine Verknüpfung zwischen dem Leistungsdichtespektrum eines Zufallsprozesses und dessen Autokorrelationsfunktion. Daher werden im anschließenden Abschnitt traditionelle Periodogramm-Verfahren zur Spektralschätzung vorgestellt. Aus dem so geschätzten Leistungsdichtespektrum lassen sich durch Rücktransformation in den Zeitbereich ebenfalls die Werte der Autokorrelationsfolge bestimmen.

Allen Methoden gemein ist, dass ihre Eingangssignale aus Blöcken von Daten der endlichen Länge L_{AR} bestehen, in denen Kurzzeit-Stationarität (siehe Abschnitt 2.1.1) angenommen werden kann.

4.2.1 Schätzung der Autokorrelationsfunktion

Die einfachste Methode, eine Autokorrelationsfolge zu schätzen, besteht in dem Verwenden ihrer Definition, wobei der Erwartungswert durch eine normierte Summe über die Zeit ersetzt wird:

$$\hat{r}_{yy}(l) \;=\; \frac{1}{L_{\mathrm{AR}}} \sum_{k=0}^{L_{\mathrm{AR}}-1} y^*(k)y(k+l). \tag{4.22}$$

Da die zur Verfügung stehenden Daten auf das Interval $0 \leq k \leq L_{\mathrm{AR}}-1$ beschränkt sind, müssen die Summationsgrenzen angepasst werden. Dazu werden die Daten außerhalb des Intervalls implizit zu Null angenommen. Mit größer werdendem l nimmt daher die Anzahl der verwendeten Summanden immer weiter ab. Im Extremfall $l = L_{\mathrm{AR}}-1$ entfällt die Summation, da nur noch ein Summand übrig bleibt. Mit den so angepassten Summationsgrenzen erhält man die sogenannte *Autokorrelationsmethode*[3]:

$$\hat{r}_{yy}(|l|) \;=\; \frac{1}{L_{\mathrm{AR}}} \sum_{k=0}^{L_{\mathrm{AR}}-1-|l|} y^*(k)y(k+|l|). \tag{4.23}$$

Die Werte für negative Verschiebungen l können aufgrund der konjugiert geraden Symmetrie der Autokorrelationsfunktion aus den positiven berechnet werden:

$$\hat{r}_{yy}(-|l|) \;=\; \hat{r}^*_{yy}(|l|). \tag{4.24}$$

Die so erhaltenen Schätzwerte für die Autokorrelationsfolge sind allerdings nicht erwartungstreu. Für den Erwartungswert der so geschätzten Autokorrelationsfunktion im Intervall $-(L_{\mathrm{AR}}-1) \leq l \leq (L_{\mathrm{AR}}-1)$ gilt:

$$\mathrm{E}\left\{\hat{r}_{yy}(|l|)\right\} \;=\; \frac{L_{\mathrm{AR}}-|l|}{L_{\mathrm{AR}}} r_{yy}(|l|). \tag{4.25}$$

Dies bedeutet, dass die wahre Autokorrelationsfolge bei diesem Schätzverfahren mit einem Dreieckfenster, dem sogenannten *Bartlett-Fenster*, gewichtet wird. Grund für die fehlende Erwartungstreue ist die ungleiche Gewichtung für verschiedene Werte der Verschiebung l. Je größer diese Verschiebung ist, desto weniger

[3] Genau genommen spricht man erst von der Autokorrelationsmethode, wenn die AR-Parameter mittels der Yule-Walker-Gleichung 4.11 berechnet werden und die dazu notwendige Autokorrelationsmatrix mit Gleichung 4.23 geschätzt wurde.

Werte gehen aufgrund der auf L_{AR} begrenzten Blocklänge in die Berechnung ein. Trotzdem werden alle Verschiebungen auf den konstanten Faktor L_{AR} normiert.

Betrachtet man Gleichung 4.25 für gegenüber der Blocklänge nur sehr kleine Verschiebungen, also für $L_{\mathrm{AR}} \gg l$, dann verschwindet für $L_{\mathrm{AR}} \to \infty$ der Einfluss des Bartlett-Fensters. Es liegt unter diesen Voraussetzungen also asymptotische Erwartungstreue vor.

Um einen nicht nur asymptotisch erwartungstreuen Schätzwert zu erhalten, muss das in Gleichung 4.23 vorgestellte Schätzverfahren modifiziert werden, so dass auf die jeweils eingehende Anzahl von Werten in der Summe normiert wird. Man erhält die sogenannte *modifizierte Autokorrelationsmethode*:

$$\hat{r}_{yy}(|l|) \;=\; \frac{1}{L_{\mathrm{AR}} - |l|} \sum_{k=0}^{L_{\mathrm{AR}}-1-|l|} y^*(k)y(k+|l|). \tag{4.26}$$

In diesem Fall kürzt sich der Faktor vor der wirklichen Autokorrelationsfunktion in Gleichung 4.25 zu Eins, und man erhält eine erwartungstreue Schätzung.

Mindestens genauso wichtig wie die Erwartungstreue eines Schätzwertes ist die Konvergenz seiner Varianz gegen Null für $L_{\mathrm{AR}} \to \infty$. Einen solchen Schätzwert nennt man *konsistent*. Dabei ist es unerheblich, ob die vorliegende Erwartungstreue nur asymptotisch ist oder nicht. Unter der Annahme, dass die beteiligten Zufallsprozesse für Real- und Imaginärteil jeweils unkorrelierte, mittelwertfreie Gaußprozesse sind, kann gezeigt werden, dass die Varianz der Schätzung nach Gleichung 4.23 für $L_{\mathrm{AR}} \to \infty$ verschwindet. Bei der modifizierten Autokorrelationsmethode nach Gleichung 4.26 muss dafür zusätzlich $L_{\mathrm{AR}} >> l$ gefordert werden, da hier Schätzwerte, die aus nur wenigen Eingangswerten gebildet wurden (l groß) genauso eingehen wie die aus vielen gebildeten (l klein) [26, 46]. Die bedeutet, dass bei der Schätzung nach Gleichung 4.26 die Fehlervarianz mit wachsendem l zunimmt. Dies ist der Preis für die Erwartungstreue.

Beide Verfahren erzeugen eine Autokorrelationsmatrix, die eine Toeplitz-Struktur aufweist. Aufgrund dieser Eigenschaft werden sie auch als stationäre Verfahren bezeichnet. Hinzu kommt, dass die so gewonnene Autokorrelationsmatrix prinzipiell immer invertierbar ist. Daher liefern sowohl die Autokorrelationsmethode als auch die modifizierte Autokorrelationsmethode stabile AR-Koeffizienten.

In den in dieser Arbeit vorgeschlagenen Verfahren kann für die Blocklänge nicht $L_{\mathrm{AR}} >> l$ angenommen werden, da aufgrund der Unterabtastung durch die Filterbankimplementierung und der gewählten Abtastfrequenz relativ kurze Datenblöcke entstehen und die maximale Verschiebung $l = l_{\max}$ im Verhältnis dazu nicht vernachlässigt werden kann. Die Länge (in Abtastwerten) der Datenblöcke hängt von der Abtastfrequenz ab, da jeder Block aufgrund der Stationaritätsannahme bei Sprache nicht mehr als ca. 50 ms Signal enthalten darf. Durch die Unterabtastung der Filterbank wird die Blocklänge dann noch einmal um den Faktor r

reduziert. Daher wird das Konzept der modifizierten Autokorrelationsmethode in den späteren Verfahren nicht verwendet.

Alternativ kann die Autokorrelationsmethode aus Gleichung 4.23 analog zur schnellen Faltung über die schnelle Fouriertransformation (FFT) realisiert werden. In diesem Zusammenhang ist insbesondere das Verfahren nach Rader [41] zu erwähnen.

4.2.2 Periodogramm-Verfahren zur Schätzung des Leistungsdichtespektrums

Eine weitere Möglichkeit, die AR-Parameter zu berechnen, besteht darin, zunächst das Leistungsdichtespektrum zu schätzen. Dieses ist über die Wiener-Khintchine-Beziehung mit der Autokorrelationsfunktion verknüpft. Somit kann durch Rücktransformation die gewünschte Autokorrelationsfolge bestimmt werden.

Die Stabilität der daraus resultierenden AR-Koeffizienten ist in diesem Fall gegeben, wenn das geschätzte Leistungsdichtespektrum für alle Frequenzen größer Null ist.

Das bekannteste dieser unter dem Begriff *nicht-parametrische Spektralschätzung* zusammengefassten Verfahren ist das Periodogramm. Dieses ist definiert als das auf die Blocklänge L_{AR} normierte Betragsquadrat eines in den Frequenzbereich transformierten Signalblocks:

$$\hat{S}_{yy}(\Omega) = \frac{1}{L_{\mathrm{AR}}} \left| Y\left(e^{j\Omega}\right) \right|^2 = \frac{1}{L_{\mathrm{AR}}} \sum_{k_1=0}^{L_{\mathrm{AR}}-1} \sum_{k_2=0}^{L_{\mathrm{AR}}-1} y^*(k_1) y(k_2) e^{-j\Omega(k_2-k_1)}. \quad (4.27)$$

Mit der Substitution $l = k_2 - k_1$ kann Gleichung 4.27 für $l \geq 0$ folgendermaßen umgeformt werden [26]:

$$\begin{aligned} \hat{S}_{yy}(\Omega) = & \sum_{l=0}^{L_{\mathrm{AR}}-1} \left[\frac{1}{L_{\mathrm{AR}}} \sum_{k=0}^{L_{\mathrm{AR}}-1-l} y^*(k) y(k+l) \right] e^{-j\Omega l} \ldots \\ & + \sum_{l=-(L_{\mathrm{AR}}-1)}^{-1} \left[\frac{1}{L_{\mathrm{AR}}} \sum_{k=-l}^{L_{\mathrm{AR}}-1} y^*(k) y(k+l) \right] e^{-j\Omega l} \end{aligned} \quad (4.28)$$

Das Periodogramm entspricht also der Fourier-Transformierten der Schätzung der Autokorrelationsfolge nach Gleichung 4.23. Genau wie diese ist das geschätzte Leistungsdichtespektrum nur asymptotisch erwartungstreu.

Betrachtet man die Varianz des Periodogramms nach Gleichung 4.28, so verschwindet diese für $L_{\mathrm{AR}} \to \infty$ nicht, sondern konvergiert gegen den konstanten

Wert σ_y^4 [32]. Damit ist das Periodogramm keine konsistente Schätzung des Leistungsdichtespektrums.

Um dennoch Konsistenz zu erreichen, müssen mehrere Periodogramme gemittelt werden. Bei diesem als Bartlett-Methode bezeichneten Verfahren wird der Datenblock der Länge L_{AR} in K nicht-überlappende Teilfolgen $y_i(k)$ der Länge L_K zerlegt. Dabei wird angenommen, dass die Periodogramme aufeinander folgender Teilstücke des Datenblocks der Länge L_{AR} unabhängig sind:

$$K = \frac{L_{\mathrm{AR}}}{L_K}, \tag{4.29}$$

Für jede Teilfolge wird nun ein Periodogramm $\hat{S}_{yy}(\Omega, i)$ für $i = 1, \ldots, K$ der Länge L_K gemäß Gleichung 4.27 berechnet. Durch Mittelung dieser K Periodogramme erhält man schließlich die Schätzung des Leistungsdichtespektrums nach Bartlett:

$$\hat{S}_{yy,\mathrm{Bartlett}}(\Omega) = \frac{1}{K} \sum_{i=1}^{K} \hat{S}_{yy}(\Omega, i). \tag{4.30}$$

Die Varianz dieser Schätzung verkleinert sich um den Faktor K. Allerdings verringert sich die spektrale Auflösung ebenfalls um den Faktor K, da weniger Punkte in der FFT berechnet werden. Für $L_{\mathrm{AR}} \to \infty$ wächst sowohl die Anzahl K als auch die Länge L_K der Teilfolgen. Während ersteres für eine kleiner werdende Varianz sorgt, verbessert letzteres die Erwartungstreue. Unter diesem Gesichtspunkt kann die Bartlett-Methode als konsistente Schätzung des Leistungsdichtespektrums verstanden werden. Die Wahl dieser Parameter bei gegebenem L_{AR} stellt also immer einen Kompromiss dar.

Das Verfahren nach Bartlett gewichtet die einzelnen Teilfolgen mit einem rechteckförmigen Fenster. Dessen Auswirkungen entsprechen denen der Spektralanalyse deterministischer Signale. Das heißt, es tritt der *Leckeffekt* auf. Multipliziert man die Teilfolgen vor Bildung des Periodogramms daher mit einer geeigneten Fensterfunktion wie zum Beispiel dem Hanning-Fenster, so kann der Einfluss dieses Effekts verringert werden. Man erhält die als Welch-Methode bezeichnete Verbesserung der Bartlett-Methode:

$$\hat{S}_{yy,\mathrm{Welch}}(\Omega) = \frac{1}{\varrho K} \sum_{i=1}^{K} \frac{1}{L_K} \left| \sum_{k=0}^{L_K-1} y_i(k) w(k) e^{-j\Omega k} \right|^2, \tag{4.31}$$

wobei $w(k)$ die gewählte Fensterfunktion der Länge L_K bezeichnet und ϱ den für die asymptotische Erwartungstreue notwendigen Normierungsfaktor darstellt. Dieser berechnet sich zu [26]:

$$\varrho = \frac{1}{L_K} \sum_{k=0}^{L_K-1} w^2(k). \tag{4.32}$$

Für beide Methoden ist es möglich, die Teilfolgen überlappend zu wählen. In praktischen Anwendungsfällen wird die Unabhängigkeit der einzelnen Periodogramme nicht gefährdet, wenn die einzelnen Teilfolgen sich nicht um mehr als die Hälfte überlappen [26]. Die Anzahl der Teilfolgen verdoppelt sich in diesem Beispiel näherungsweise auf:

$$K_{50\%\text{overlap}} \quad = \quad 2\frac{L_{\text{AR}}}{L_K} - 1, \tag{4.33}$$

wodurch sich die Varianz ebenfalls nahezu halbiert.

Neben den Periodogramm basierten Methoden gibt es eine weitere oft verwendete Klasse von Schätzalgorithmen, die auf dem *Korrelogramm* oder auch der *Blackman-Tukey-Schätzung* basieren. Diese Verfahren erzeugen allerdings keine garantiert positiven Leistungsdichtespektren [26]. Somit ist die Stabilität der damit geschätzten AR-Modelle nicht sichergestellt, weshalb diese Klasse nicht näher untersucht wird.

4.2.3 Methoden zur Berechnung der AR-Parameter

In Abschnitt 4.1.2 wurde das Problem der AR-Parameter Schätzung auf die Berechnung eines Schätzwertes der Autokorrelationsfolge reduziert. Daraufhin wurden in Abschnitt 4.2 sowohl Zeit- als auch Frequenzbereichsmethoden vorgestellt, die eine solche Schätzung der Autokorrelationsfolge ermöglichen. Damit fehlt nur noch die eigentliche Berechnung der AR-Parameter ausgehend von dieser Schätzung. Dies wird im folgenden Abschnitt behandelt.

Um zu den gesuchten AR-Parametern $a_i(k)$ für $i = 0, \ldots, L_{\text{AR}} - 1$ zu gelangen, gibt es zwei mögliche Vorgehensweisen:

- Direkte Methoden: Diese basieren auf der Lösung des Normalen-Gleichungssystems aus Gleichung 4.8. Dies entspricht dem Ansatz eines transversalen Prädiktorfehlerfilters.
- Rekursive Methoden: Diese berechnen die sogenannten *Reflexionskoeffizienten*, welche eine äquivalente Beschreibung des AR-Modells darstellen. Dies entspricht dem Ansatz eines Prädiktorfehlerfilters in *Lattice*-Struktur[4] (siehe Abbildung 4.2).

In dieser Arbeit wird sowohl eine direkte wie auch eine rekursive Methode verwendet. Diese werden im Folgenden kurz vorgestellt. Die Ordnung des AR-Modells sei dabei p.

[4] Für den genauen Zusammenhang zwischen Reflexionskoeffizienten und Lattice-Struktur beim linearen Prädiktor siehe [19].

Eigenschaften der Reflexionskoeffizienten

In diesem Abschnitt soll kurz auf die Eigenschaften der Reflexionskoeffizienten, welche mit $\Gamma_i(k)$ für $i=1,\ldots,p$ notiert werden, eingegangen werden.

Wie bereits erwähnt, sind die Reflexionskoeffizienten eine zu den AR-Koeffizienten äquivalente Beschreibung des AR-Modells. Es macht also keinen Unterschied, ob das Modell mit p Koeffizienten $a_i(k)$ oder p Koeffizienten $\Gamma_i(k)$ beschrieben wird. Für eine vollständige Beschreibung fehlt natürlich noch jeweils die Anregungsleistung σ_v^2.

Zwei Unterschiede sind allerdings bemerkenswert:

1. Im Gegensatz zu den AR-Koeffizienten, für die $a_i(k) \in \mathbb{R}$ gilt, ist der Wertebereich der Reflexionskoeffizienten $\Gamma_i(k)$ beschränkt, solange das AR-Modell stabil ist. In diesem Fall gilt: $|\Gamma_i(k)| < 1$.

2. Reflexionskoeffizienten verändern sich nicht bei Erhöhung der Ordnung. Das heißt, beim Übergang von Ordnung p auf $p+1$ kommt nur der neue Koeffizient hinzu. Alle vorhanden bleiben unverändert. Im Gegensatz dazu müssen bei der Darstellung mittels AR-Koeffizienten alle bereits vorhanden umgerechnet – also verändert – werden.

Besonders die erste Eigenschaft hat viele praktische Auswirkungen. So kann beispielsweise die Stabilität der geschätzten $\Gamma_i(k)$ leicht überprüft werden, indem folgende Bedingung geprüft wird:

$$|\Gamma_i(k)| < 1 \quad \text{für} \quad i = 1,\ldots,p. \tag{4.34}$$

Außerdem bieten die Reflexionskoeffizienten eine Möglichkeit, verschiedene Sätze von Koeffizienten zu einem Satz zu vereinen. Dies wird später benötigt, wenn die AR-Parameter für das Sprachmodell in mehreren Kanälen unabhängig voneinander geschätzt und zu einem gemittelten Satz zusammengefügt werden sollen. Dies ist in der Darstellung mit AR-Koeffizienten schwierig. Im Gegensatz dazu kann bei der Addition mehrere Sätze stabiler Reflexionskoeffizienten sofort mittels Gleichung 4.34 auf Stabilität geprüft werden. Zum Beispiel gilt für das arithmetische Mittel von stabilen Reflexionskoeffizientensätzen:

$$\overline{\Gamma_i(k)} = \frac{1}{K}\left(\Gamma_i^{(1)}(k) + \Gamma_i^{(2)}(k) + \cdots + \Gamma_i^{(K)}(k)\right) \quad \text{für} \quad i = 1,\ldots,p. \tag{4.35}$$

Die Normierung der Summe auf die Anzahl K der aufaddierten Sätze bewirkt hier, dass die gemittelten Koeffizienten $\overline{\Gamma_i(k)}$ ebenfalls dem Betrage nach kleiner als Eins sind, was die Stabilität des zugehörigen AR-Modells garantiert. Das arithmetische Mittel ist hier ein Sonderfall der gewichteten Addition mit anschließender Normierung, bei der jeder Satz unterschiedlich gewichtet werden kann. Bei

der Normierung der Summe muss lediglich sichergestellt werden, dass $|\overline{\Gamma_i(k)}| < 1$ gilt.

Aufgrund dieser Eigenschaften ist es in vielen Fällen geschickter, mit den Reflexionskoeffizienten $\Gamma_i(k)$ zu rechnen, insbesondere da mittels Step-Up- bzw. Step-Down-Algorithmus leicht zwischen beiden Darstellungen umgerechnet werden kann.

Direkte Methoden zur Berechnung der AR-Parameter

Entsprechend der Definition der direkten Methoden, wird von der Lösung des Normalen-Gleichungssystems aus Gleichung 4.11 ausgegangen. Die zu invertierende Autokorrelationsmatrix $\hat{\mathbf{R}}_{yy}(k)$ enthält jetzt aber Schätzwerte der Autokorrelationsfolge, weshalb sie ebenfalls mit dem Dachoperator notiert wird. Für deren Schätzung verwendet man die Autokorrelationsmethode gemäß Gleichung 4.23 aus Abschnitt 4.2, wodurch ein stabiles AR-Modell garantiert ist und $\hat{\mathbf{R}}_{yy}(k)$ zudem eine konjugiert-symmetrische Toeplitz-Struktur aufweist.

Eine Inversion von $\hat{\mathbf{R}}_{yy}(k)$ mit dem Gaußverfahren bedeutet einen numerischen Aufwand von $\sim p^3$. Unter Ausnutzung der besonderen Struktur der Matrix (Toeplitz), kann der Aufwand um eine Größenordnung auf $\sim p^2$ gesenkt werden. Ein Verfahren, das dies leistet, ist beispielsweise der *Levinson-Durbin-Algorithmus*. Da Teile von diesem außerdem zur Umrechnung zwischen Reflexions- und AR-Koeffizienten verwendet werden, wird dessen Funktionsweise hier kurz erläutert. Für eine detaillierte Beschreibung und Herleitung sei auf [19, 21, 26, 46] verwiesen.

Der Levinson-Durbin-Algorithmus invertiert die zuvor geschätzte Autokorrelationsmatrix nicht direkt, sondern mittels einer Rekursion über die Ordnung. Dies hat allerdings nichts mit der Unterteilung in direkte und rekursive Methoden zu tun. Am Beispiel der Ordnung p wird in drei Schritten folgende Berechnung durchgeführt:

1. Berechnung des sogenannten *PARCOR*-Koeffizienten[5] der Ordnung p aus den Daten der geschätzten Autokorrelationsfolge der Ordnung p sowie den bereits gefunden AR-Koeffizienten der Ordnung $(p-1)$. Die Umrechnung von Autokorrelationswerten in Reflexionskoeffizienten wird auch als *Schur-Rekursion* bezeichnet.

2. Berechnung der Anregungsleistung der Ordnung p aus derjenigen der Ordnung $(p-1)$ unter Verwendung des PARCOR-Koeffizienten.

[5] Der PARCOR-Koeffizient der Ordnug p entspricht dem Reflexionskoeffizienten $\Gamma_p(k)$ der entsprechenden Ordnung.

3. Berechnung der AR-Koeffizienten der Ordnung p aus denen der Ordnung $(p-1)$ mittels des PARCOR-Koeffizienten der Ordnung p. Dieser Teil wird auch als *Step-Up-Algorithmus* bezeichnet, seine Umkehrung – also von AR-Koeffizienten zu Reflexionskoeffizienten – als *Step-Down-Algorithmus*.

Alternativ zu Schritt zwei kann die Anregungsleistung auch nach der Berechnung der AR-Koeffizienten gemäß Gleichung 4.12 bestimmt werden.

Die hier vorgestellte direkte Methode hat den Nachteil, dass aufgrund der nur asymptotisch erwartungstreuen Schätzung der Autokorrelationsfolge gemäß Gleichung 4.23 ein Fehler entsteht, da die Werte außerhalb des Blockintervalls implizit zu Null angenommen werden. Ein Verfahren, das dies vermeidet ist die sogenannte *Kovarianzmethode*. Da dieses aber im Gegensatz zur Autokorrelationsmethode keine stabilen AR-Modelle garantiert [19, 26, 46], wird sie hier nicht näher betrachtet.

Rekursive Methoden zur Berechnung der AR-Parameter

Am Ende des letzten Abschnitts wurde bemerkt, dass die hier genannten direkten Methoden den Nachteil haben, dass sie entweder eine erwartungstreue Schätzung oder aber stabile AR-Modelle liefern. Wünschenswert ist ein Verfahren, dass beide Voraussetzungen erfüllt. Dies leistet der *Burg-Algorithmus*, welcher zu den rekursiven Verfahren zählt.

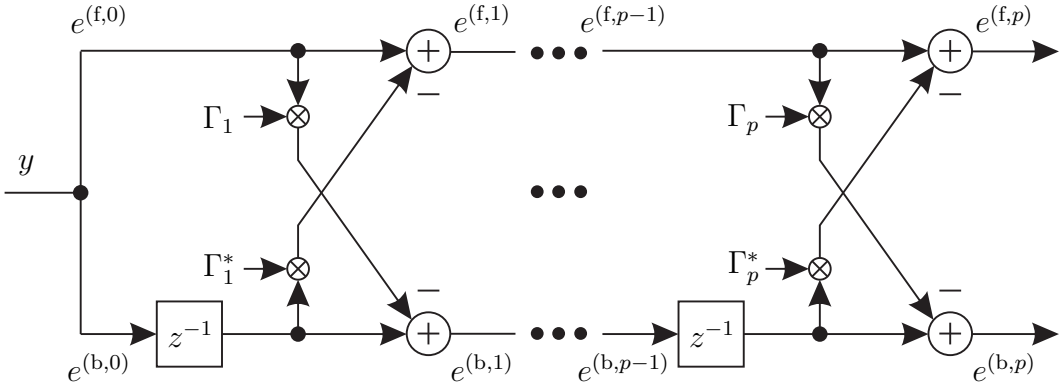

Abbildung 4.2: *Die von rekursiven Verfahren verwendete Lattice Struktur. Aus Gründen der Übersichtlichkeit wurden alle Größen ohne Zeitindex k notiert.*

Die rekursiven Methoden nutzen eine Kaskadierung von Teilsystemen, um die AR-Koeffizienten $a_i(k)$ schlussendlich aus den Reflexionskoeffizienten zu bestimmen. Jedes dieser Teilsystem liefert den Reflexionskoeffizient der nächst höheren Ordnung angefangen mit dem der Ordnung Eins. Die Anpassung des Modells

an das vorliegende Signal $y(k)$ liefert hier also einen Satz Reflexionskoeffizienten. Am Ende der Rekursion müssen diese daher mittels Step-Up-Algorithmus in AR-Koeffizienten umgerechnet werden. Dieses als *Lattice-Struktur* bezeichnete System kaskadierter Teilsysteme ist in Abbildung 4.2 dargestellt, wobei $e^{(\mathrm{f},p)}$ und $e^{(\mathrm{b},p)}$ den Vorwärts- bzw. Rückwärts-Prädiktionsfehler bezeichnen.

Im Gegensatz zu den bereits beschriebenen Verfahren, minimiert der Burg-Algorithmus sowohl den mittleren quadratischen Vorwärts- als auch den Rückwärts-Prädiktionsfehler. Die Reflexionskoeffizienten des Modells werden rekursiv über die Ordnung berechnet, wobei für den Reflexionskoeffizient der Ordnung p zum Zeitpunkt k gilt [21, 26, 46]:

$$\Gamma_p(k) \;=\; \frac{2\sum_{k=p}^{L_{\mathrm{AR}}-1}\left(e^{(\mathrm{f},p-1)}(k)\cdot \mathrm{conj}\{e^{(\mathrm{b},p-1)}(k-1)\}\right)}{\sum_{k=p}^{L_{\mathrm{AR}}-1}\left(\left|e^{(\mathrm{f},p-1)}(k)\right|^2+\left|e^{(\mathrm{b},p-1)}(k-1)\right|^2\right)}. \tag{4.36}$$

Initialisiert wird das Verfahren mit:

$$e^{(\mathrm{f},0)}(k)=e^{(\mathrm{b},0)}(k)=y(k) \quad \text{für} \quad k=0,\ldots,L_{\mathrm{AR}}-1. \tag{4.37}$$

Wie im vorangegangenen Abschnitt bereits erwähnt, ist ein AR-Modell genau dann stabil, wenn die zugehörigen Reflexionskoeffizienten Gleichung 4.34 genügen. Um dies zu zeigen, wird Gleichung 4.36 in Vektorschreibweise übersetzt. Unter Vernachlässigung des Zeitindex' k ergibt sich dann:

$$|\Gamma_p| \;=\; \frac{2\left|\left(\mathbf{e}^{(\mathrm{f},p-1)}\right)^{\mathrm{T}}\mathrm{conj}\left\{\mathbf{e}^{(\mathrm{r},p-1)}\right\}\right|}{\left\|\mathbf{e}^{(\mathrm{f},p-1)}\right\|^2+\left\|\mathbf{e}^{(\mathrm{r},p-1)}\right\|^2}. \tag{4.38}$$

Die Dreiecksungleichung [8] besagt, dass für zwei beliebige, komplexwertige Vektoren $\mathbf{x}$ und $\mathbf{y}$ gilt:

$$2\left|\mathbf{x}^{\mathrm{H}}\mathbf{y}\right| \;\leq\; \|\mathbf{x}\|^2+\|\mathbf{y}\|^2. \tag{4.39}$$

Ein Vergleich von Gleichung 4.38 und 4.39 liefert sofort:

$$|\Gamma_p| \;\leq\; 1, \tag{4.40}$$

wobei für Gleichheit beide Vektoren identisch sein müssen. Die mit dem Burg-Algorithmus erhaltene Schätzung garantiert somit stabile AR-Modelle. Der Preis dafür ist der höhere numerische Aufwand und die zusätzlich erforderliche Umrechnung von Reflexions- in AR-Koeffizienten. Da in den vorgeschlagenen Verfahren jeweils die Reflexionskoeffizienten zur weiteren Verarbeitung benötigt werden, ist dies hier nicht als Nachteil zu werten.

4.3 Mehrkanalige Schätzung der AR-Parameter

Im nun folgenden Abschnitt werden die bisher gefunden Methoden kombiniert, um die im Kalman-Filter aus Kapitel 3 benötigten AR-Parameter zu schätzen. Hierbei ist zu beachten, dass dazu sowohl die AR-Parameter des ungestörten Sprachsignals als auch die der N Rauschsignale bestimmt werden müssen. Dies bedeutet, dass bereits vor der eigentlichen Geräuschreduktion durch das Kalman-Filter eine Trennung von Sprache und Geräusch zur Schätzung der jeweiligen AR-Parameter durchgeführt werden muss. Die Güte dieser Trennung hat daher großen Einfluss auf die Leistungsfähigkeit des Gesamtsystems.

Für die einkanalige Methode in [39] wurde hierzu ein *Differenz-Autokorrelationsmethode* (abgekürzt: DAKF-Methode) genanntes Verfahren vorgestellt. Im Folgenden werden aufbauend auf diesem verschiedene mehrkanalige Erweiterungen vorgeschlagen.

Im vorangegangenen Abschnitt wurde der Zusammenhang zwischen Zeit- und Spektralbereich stets über die *zeitdiskrete Fourier-Transformation* (DTFT) beschrieben. Dieses mathematische Konstrukt liefert im Frequenzbereich eine kontinuierliche Funktion der normierten Frequenz Ω. Für praktische Anwendungen mit digitalen Rechnern müssen aber sowohl der Zeit- als auch der Spektralbereich durch diskrete Größen beschrieben werden. Daher wird im Folgenden Abschnitt die DTFT durch die *diskrete Fouriertransformation* (DFT) und ihre Inverse (IDFT) ersetzt, wobei für die algorithmische Umsetzung ausnahmslos die schnelle Fouriertransformation (FFT) und ihre Umkehrung (IFFT) [32] verwendet werden. Aufgrund der blockweisen Verarbeitung wird zur eindeutigen Bezeichnung neben dem Frequenzindex n auch der Zeitindex k bei allen Frequenzbereichsgrößen und Autokorrelationsfolgen mitgeführt. Durch die Verwendung des in der deutschen Literatur gebräuchlichen Frequenzindex' n ergibt sich eine Doppeldeutigkeit mit dem Geräuschsignal $n(k)$ bzw. dessen DFT-Transformierten $N_k(n)$. Aus dem Zusammenhang ist allerdings stets eindeutig zu ersehen, ob mit n der Frequenzindex oder das Geräuschsignal gemeint ist. Unter Verwendung dieser Notation gilt für den Zusammenhang zwischen Autokorrelationsfolge und Leistungsdichtespektrum:

$$\hat{r}_{yy,k}(l) \;=\; \mathrm{IDFT}\left\{\hat{S}_{yy,k}(n)\right\}. \tag{4.41}$$

4.3.1 DAKF-Methode im einkanaligen Fall

Die Funktionsweise der DAKF-Methode ist in Abbildung 4.3 dargestellt. Der Name rührt von der Kombination der Autokorrelationsmethode und der *spektralen Subtraktion* [53] her.

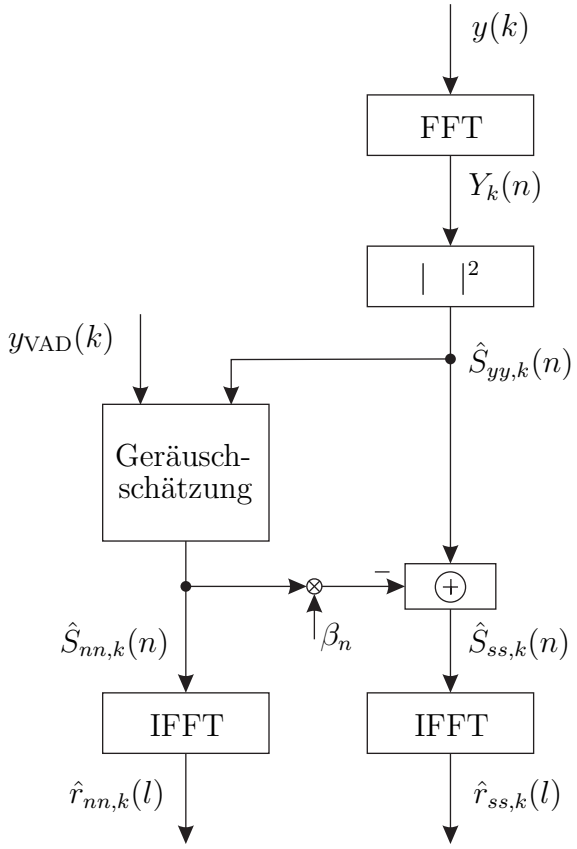

Abbildung 4.3: *Signalflussdiagramm der einkanaligen DAKF-Methode nach [39]. Der Kasten um das Summationszeichen bedeutet, dass hier negative Werte abgefangen werden.*

Ausgehend von einem komplexwertigen Datenblock $y(k)$ der Länge L_{AR} wird zunächst das Periodogramm $\hat{S}_{yy,k}(n)$ durch Transformation mittels DFT und Bildung des Betragsquadrats analog zu Gleichung 4.27 berechnet. Da das Periodogramm $\hat{S}_{yy,k}(n)$ eine Kurzzeit-Schätzung des Leistungsdichtespektrums von $y(k)$ darstellt, wird es mit dem Dach-Operator notiert:

$$\hat{S}_{yy,k}(n) \;=\; \frac{1}{L_{\mathrm{AR}}} \left|\mathrm{DFT}\left\{y(k)\right\}\right|^2 . \tag{4.42}$$

Um zyklische Effekte durch die Verwendung der DFT zu vermeiden, muss die Folge $y(k)$ vor dem Ausführen von Gleichung 4.42 durch Anhängen von Nullen (Englisch: Zero-Padding) verlängert werden. Die Anzahl der anzufügenden Nullen richtet sich nach der Ordnung des AR-Modells. Da hier gleichzeitig die Parameter mehrerer AR-Modelle geschätzt werden, das sind das Sprachmodell der Ordnung p sowie die Geräuschmodelle der Ordnung q, muss die größte dieser Ordnungen verwendet werden. In der vorliegenden Arbeit gilt stets $p \geq q$, weshalb der Da-

tenblock um mindestens p Nullen verlängert werden muss [40]. Dies rührt daher, dass für die Berechnung der AR-Parameter der Ordnung p die ersten $p+1$ Werte der Autokorrelationsfolge $r_{yy,k}(l)$, das sind die Autokorrelationswerte für die Verschiebungen $l=0,\ldots,p$, vorhanden sein müssen. Ohne Zero-Padding ist nur der Wert für $l=0$ korrekt, während alle anderen aufgrund der zyklischen Faltung verfälscht sind. Für jede weitere Null erhöht sich die Anzahl der Werte, die denen durch lineare Faltung berechneten entsprechen, um eins. Die dabei entstehende neue Blocklänge wird mit L'_{AR} bezeichnet:

$$L'_{\mathrm{AR}} \geq L_{\mathrm{AR}} + p \quad \text{mit} \quad 0 < p \leq L_{\mathrm{AR}}. \tag{4.43}$$

Für eine schnelle Implementierung der DFT wird der FFT-Algorithmus [32] verwendet. Daher ist es sinnvoll, für die Blocklänge L'_{AR} eine Zweierpotenz zu wählen. Das gesamte Vorgehen entspricht der *schnellen Faltung* der zeitdiskreten Signalverarbeitung [32].

In dem Block Geräuschschätzung wird das Leistungsdichtespektrum $\hat{S}_{nn,k}(n)$ des Geräuschs $n(k)$ aus dem Summenspektrum $\hat{S}_{yy,k}(n)$ geschätzt. Hierzu ist eine Detektion der Sprachpausen (VAD) notwendig, deren Untersuchung nicht Teil dieser Arbeit ist[6]. Diese wird durch das Signal $y_{\mathrm{VAD}}(k)$ abgebildet, welches folgendermaßen definiert ist:

$$y_{\mathrm{VAD}}(k) \;=\; \begin{cases} 1 & \text{bei Sprachaktivität} \\ 0 & \text{in Sprachpausen} \end{cases} . \tag{4.44}$$

Für die Erklärung der DAKF-Methode wird angenommen, dass die Sprachpausendetektion fehlerfrei funktioniert. Das Signal $y_{\mathrm{VAD}}(k)$ wird daher ohne den Dachoperator geschrieben.

Liegt zum Zeitpunkt k eine Sprachpause vor, so enthält $\hat{S}_{yy,k}(n)$ nur Geräuschanteile und es gilt: $\hat{S}_{nn,k}(n) = \hat{S}_{yy,k}(n)$. Man erhält in diesem Fall eine ungestörte Schätzung des Geräuschs.

Im umgekehrten Fall (für $y_{\mathrm{VAD}}(k)=1$) wird prinzipiell die Geräuschschätzung der letzten Sprachpause verwendet. Damit diese zumindest einigermaßen repräsentativ für das aktuelle Geräusch ist, muss sie während der Sprachpausen geglättet werden. Dazu wird üblicherweise ein rekursives Filter erster Ordnung verwendet, so dass für das geschätzte Leistungsdichtespektrum der Störung gilt:

$$\hat{S}_{nn,k}(n) \;=\; \begin{cases} \hat{S}_{nn,k-1}(n) & \text{für } y_{\mathrm{VAD}}(k) = 1 \\ \alpha_n \hat{S}_{nn,k-1}(n) + (1-\alpha_n)\hat{S}_{yy,k}(n) & \text{für } y_{\mathrm{VAD}}(k) = 0 \end{cases} , \tag{4.45}$$

wobei α_n die Glättungskonstante bezeichnet. Diese liegt üblicherweise im Bereich zwischen 0,9 und 0,995. Außerdem ist es möglich, für ansteigende und abfallende Signalpegel unterschiedliche Glättungskonstanten zu wählen. Dadurch wird

[6] Für den interessierten Leser sei beispielsweise auf [39, 53, 56] verwiesen.

das Verfahren robuster gegenüber Fehldetektionen des Sprachpausendetektors. Will man beispielsweise verhindern, dass in einer fälschlich erkannten Pause der Sprachanteil des Summensignals das Leistungsdichtespektrum des Rauschens zu schnell verfälscht, dann wählt man α_n für steigende Signalpegel größer als für fallende. Dadurch werden ansteigende Pegel stärker geglättet, wodurch diese deutlich länger anliegen müssen, um den gleichen Effekt auf die Geräuschschätzung zu erzielen wie fallenden Pegel [48].

Um einen Schätzwert $\hat{S}_{ss,k}(n)$ für das Leistungsdichtespektrum des Sprachanteils zu erhalten, muss das so geschätzte Geräuschspektrum vom Gesamtspektrum subtrahiert werden. Wird diese Subtraktion gemäß

$$\hat{S}_{yy,k}(n) - \hat{S}_{nn,k}(n) \tag{4.46}$$

ausgeführt, so treten dabei die gleichen Probleme wie bei der direkten Geräuschreduktion mittels spektraler Subtraktion auf. Durch die starke Glättung folgt die Geräuschschätzung nur sehr langsam den Änderungen des echten Geräuschs, während das Summensignal ungeglättet verwendet wird. Eine Subtraktion dieser Signale kann negative Werte im Leistungsdichtespektrum $\hat{S}_{ss,k}(n)$ generieren, welche abgefangen werden müssen, um instabile AR-Koeffizienten zu vermeiden. Im einfachsten Fall wird dazu nach der Subtraktion der Betrag gebildet:

$$\left|\hat{S}_{yy,k}(n) - \hat{S}_{nn,k}(n)\right|. \tag{4.47}$$

Durch diese notwendige Maßnahme und die unterschiedlich starke Glättung entstehen allerdings sogenannte *Musical Tones* [18].

Diese können durch Überschätzung des Geräuschs um den Verstärkungsfaktor $\beta_n > 1$ vermindert werden. Dies stellt die dritte Möglichkeiten dar, die Robustheit des Verfahrens durch Überschätzung des Geräuschs zu verbessern. Die anderen beiden Möglichkeiten wurden bereits in Kapitel 3 beschrieben. Wählt man β_n zu groß, wird die geschätzte Sprachkomponente verzerrt, während zu kleine Werte keine Verbesserung bezüglich der Musical Tones ergeben. In der Praxis haben sich Werte im Bereich zwischen 1,5 bis 2 als günstig erwiesen [39]. Gleichung 4.47 wird daher entsprechend erweitert und man erhält schließlich das geschätzte Leistungsdichtespektrum der Sprachkomponente:

$$\hat{S}_{ss,k}(n) \;=\; \left|\hat{S}_{yy,k}(n) - \beta_n \hat{S}_{nn,k}(n)\right|. \tag{4.48}$$

Noch bessere Ergebnisse erhält man, wenn man eine vollständige oder aber auch annähernd vollständige Reduktion des Signals $\hat{S}_{ss,k}(n)$ an den einzelnen Frequenzstützstellen n verhindert und statt dessen grundsätzlich etwas Rauschen durchlässt. Diese als *Rauschteppich* (Englisch: *Spectral Floor*) bezeichnete Methode ermöglicht es, die bereits durch die Geräuschüberschätzung reduzierten

Musical Tones aufgrund der psychoakustischen Funktionsweise des menschlichen Gehörs im durchgelassenen Rauschen zu verstecken [60].

Die oben beschriebenen Maßnahmen zur Absicherung gegen negative Werte im Leistungsdichtespektrum $\hat{S}_{ss,k}(n)$ sind in Abbildung 4.3 durch den Kasten, der die Summationsstelle einrahmt, dargestellt.

Im nächsten Schritt werden sowohl $\hat{S}_{ss,k}(n)$ als auch $\hat{S}_{nn,k}(n)$ mittels IFFT in den Zeitbereich zurücktransformiert. Man erhält die Kurzzeitschätzwerte für die Autokorrelationsfolgen der Sprachkomponente $\hat{r}_{ss,k}(l)$ und des Geräuschs $\hat{r}_{nn,k}(l)$:

$$\hat{r}_{ss,k}(l) = \text{IFFT}\left\{\hat{S}_{ss,k}(n)\right\} \quad \text{und} \tag{4.49}$$

$$\hat{r}_{nn,k}(l) = \text{IFFT}\left\{\hat{S}_{nn,k}(n)\right\}. \tag{4.50}$$

Dabei ist zu beachten, dass alle Werte für Verschiebungen außerhalb des Intervalls $0 < |l| \leq p$ durch den Einfluss der zyklischen Faltung verfälscht sind, wenn Gleichung 4.43 mit Gleichheit erfüllt wurde, was bedeutet, dass das Eingangssignal nur um genau p Nullen verlängert wurde.

4.3.2 Mehrkanalige Erweiterungen der DAKF-Methode

In diesem Teil des Kapitels werden nunmehr drei mehrkanalige Methoden zur Schätzung der AR-Parameter vorgestellt, die auf den Verfahren der vorangegangenen Abschnitte 4.2 sowie 4.3.1 basieren. Diese werden im Folgenden als *Methode 1*, *Methode 2* und *Methode 3* bezeichnet.

Ein Vergleich des in Kapitel 3 vorgestellten Modells für den einkanaligen ($N=1$) und mehrkanaligen Fall ($N>1$) zeigt, dass immer nur die Parameter einer Sprachquelle, aber die von N Geräuschquellen geschätzt werden müssen. Das heißt, dass für die Schätzung des Sprachanteils die mehrkanaligen Eingangsdaten in irgend einer Weise gemittelt werden müssen, worin der Gewinn der mehrkanaligen Verfahren liegt. Dabei bleibt die Geräuschschätzung in jedem Kanal unverändert zu der in Abschnitt 4.3.1 vorgestellten DAKF-Methode. Gleiches gilt für die Berechnung der Anregungsleistung, welche stets gemäß Gleichung 4.12 vorgenommen wird.

Für die Mittelung der Daten können verschiedene Strategien angewendet werden, je nachdem, an welcher Stelle im Signalflussdiagramm aus Abbildung 4.3 die Mittelung durchgeführt wird. Ziel muss dabei stets eine Verbesserung der Trennung von Sprach- und Geräusch-AR-Parametern sein. Eine Mittelung der Eingangsdaten $y_i(k)$ für $i=1,\ldots,N$ vor der spektralen Subtraktion ist deshalb wenig sinnvoll. Daher wird als Ansatz der im Folgenden vorgeschlagenen Methoden stets die DAKF-Methode aus dem letzten Abschnitt herangezogen.

Die Mittelung über verschiedene Mikrofone bedeutet im Allgemeinen, dass Signale unterschiedlicher Laufzeit (Mund - Mikrofon) auftreten. In einem normalen PKW beträgt der relative Laufzeitunterschied zwischen zwei weit entfernten Mikrofonen aufgrund der kleinen Raummaße, der vorliegenden Mikrofonanordnung und der Unterabtastung in der verwendeten Filterbankstruktur (siehe Abschnitt 5.1.3) weniger als ein Abtastintervall. Gegenüber der gewählten Blocklänge, die wegen der Stationaritätsannahme im Bereich von 20 bis 50 ms (siehe Abschnitt 2.1.1) liegt, kann der oben beschriebene Laufzeitunterschied vernachlässigt und somit auf einen Laufzeitausgleich an dieser Stelle verzichtet werden.

Für eine Abschätzung des numerischen Aufwands werden wieder die Anzahl der verwendeten Multiplikationen pro Teilband verwendet. Allerdings beschränken sich die Betrachtungen in diesem Teil auf die Berechnung der Reflexionskoeffizienten von Sprache sowie relative Vergleiche zwischen den einzelnen Methoden und dem einkanaligen Fall.

Methode 1

Im einfachsten Fall wird die DAKF für jeden der N Kanäle berechnet. Man erhält somit genau einen Satz AR-Parameter für jede der N Geräuschquellen und N Sätze für die (eine) Sprachquelle. Letztere müssen zu einem Satz zusammengefügt bzw. gemittelt werden. Das als *Methode 1* bezeichnetes Schätzverfahren ist in Abbildung 4.4 dargestellt.

Die Mikrofonsignale $y_i(k)$ mit $i = 1, \ldots, N$ werden dem in Abschnitt 4.3.1 vorgestellten DAKF-Algorithmus zugeführt. Wie dort bereits erwähnt wurde, benötigt das Verfahren zusätzlich eine Sprachpausendetektion. In [56] wurden mehrkanalige Erweiterungen verschiedener einkanaliger Sprachpausendetektoren beschrieben und bewertet. Dabei wurde festgestellt, dass die Detektionswahrscheinlichkeit des VADs bei Konditionen ähnlich den hier vorliegenden nur unwesentlich durch eine Hinzunahme weiterer Kanäle verbessert werden kann. Daher wird für Generierung des Signals der Sprachpausendetektion $y_{\mathrm{VAD}}(k)$ in dieser Arbeit nur ein Kanal ausgewertet. Dieser Kanal kann allerdings auf Grundlage der Schätzwerte für die Signal-zu-Geräusch Verhältnisse (SNR) aller Kanäle gewählt werden, welche aus den Ausgangssignalen der jeweiligen DAKF bestimmt werden:

$$\mathrm{SNR}_i(k) = \frac{\hat{r}_{s_i s_i,k}(0)}{\hat{r}_{n_i n_i,k}(0)} \quad \text{für} \quad i = 1, \ldots, N. \tag{4.51}$$

Unabhängig davon wird für jedes Teilband der Filterbankstruktur eine eigene Sprachpausendetektion durchgeführt, da aufgrund der spektralen Struktur von Sprache in einem Band Sprachsignalleistung vorhanden sein kann und gleichzeitig in einem anderen nicht.

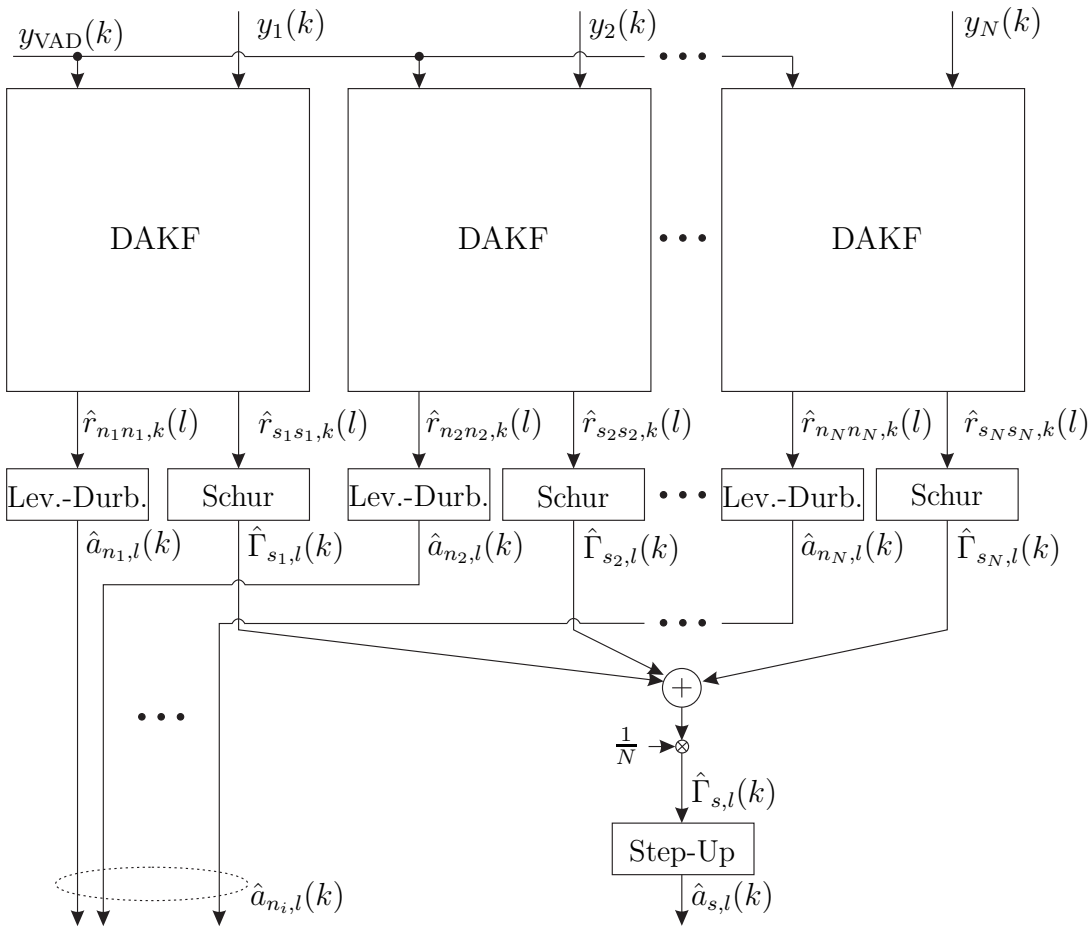

Abbildung 4.4: *Mehrkanalige* Methode 1 *zur Schätzung der AR-Parameter. Die Mittelung wird hier auf Grundlage der in Reflexionskoeffizienten* $\hat{\Gamma}_{s_i,l}(k)$ *umgerechneten Schätzwerte der Autokorrelationssequenzen* $\hat{r}_{s_is_i,k}(l)$ *der Sprachkomponente durchgeführt.*

Bei dem so bestimmten $SNR_i(k)$ muss beachtet werden, dass es erst am Ende der DAKF-Verarbeitung vorliegt und somit nur für die Entscheidung im nächsten Verarbeitungszyklus verwendet werden kann. Das heißt, dass für die Sprachpausendetektion $y_{\mathrm{VAD}}(k)$ jeweils der Kanal gewählt wird, der zum Zeitpunkt $k-1$ das maximale SNR aufgewiesen hat. Darüber hinaus können die für einen konstanten Zeitpunkt k in jedem Teilband berechneten $\mathrm{SNR}_i(k)$ unterschiedliche Werte annehmen. Die SNRs sollten daher immer innerhalb eines Teilbands verglichen werden.

Die AR-Parameter der einzelnen Geräuschquellen werden – wie im einkanaligen Fall – direkt aus den jeweiligen Schätzwerten der Autokorrelationsfolgen $\hat{r}_{n_in_i,k}(l)$ mittels Levinson-Durbin-Algorithmus bestimmt. Im Gegensatz dazu werden die Autokorrelationswerte des Sprachmodells $\hat{r}_{s_is_i,k}(l)$ mit dem Schur-Algorithmus zu

Reflexionskoeffizienten $\hat{\Gamma}_{s_i,l}(k)$ umgerechnet. Für die einzelnen Sätze gilt:

$$\left|\hat{\Gamma}_{s_i,l}(k)\right| \;<\; 1, \tag{4.52}$$

da die DAKF-Methode stets stabile AR-Koeffizienten liefert. Um den gemittelten Satz zu erhalten, werden daher die berechneten Sätze von Reflexionskoeffizienten $\hat{\Gamma}_{s_i,l}(k)$ gemäß Gleichung 4.35 aufsummiert und auf deren Anzahl normiert:

$$\hat{\Gamma}_{s,l}(k) \;=\; \frac{1}{N}\sum_{i=1}^{N}\hat{\Gamma}_{s_i,l}(k). \tag{4.53}$$

Schlussendlich wird der so erhaltene gemittelte Satz $\hat{\Gamma}_{s,l}(k)$ mittels Step-Up-Algorithmus zu AR-Koeffizienten umgerechnet.

Verglichen zum einkanaligen Fall müssen hier N DAKF- und Schur-Algorithmen berechnet werden, was dem N-fachen Aufwand entspricht. Der Vollständigkeit halber kommt eine weitere Multiplikation mit dem Normierungsfaktor $1/N$ hinzu, welche allerdings gegenüber dem Gesamtaufwand vernachlässigt werden kann.

Methode 2

Um die zuvor beschriebene Mittelung der Reflexionskoeffizienten zu verbessern, können die einzelnen Sätze vor der Addition mit dem SNR des jeweiligen Kanals (berechnet gemäß Gleichung 4.51) gewichtet werden. Dieses Gewichtungsschema ist in Abbildung 4.5 dargestellt. Die Vorverarbeitung entspricht dabei derjenigen der *Methode 1* (siehe Abbildung 4.4), weshalb sie nicht nochmals abgebildet wurde.

Um die Stabilität des resultierenden Sprach-AR-Modells zu gewährleisten, müssen die aufsummierten Reflexionskoeffizienten wieder normiert werden. Aufgrund der durchgeführten Gewichtung wird zur Normierung diesmal die Summe der einzelnen SNR-Werte $\sum_{i=1}^{N}\mathrm{SNR}_i(k)$ verwendet.

Ein relativer Vergleich beider Methoden ergibt, dass hier zusätzlich N Multiplikationen für die Normierung erforderlich sind. Der Normierungsfaktor ist jetzt zeitinvariant, so dass eine echte Division, die je nach Mikroprozessor etwa dem vierfachen Aufwand einer Multiplikation entspricht, erforderlich wird. Der verursachte Mehraufwand bezüglich *Methode 1* kann demnach als gering bezeichnet werden.

Methode 3

Bei dieser Methode setzt die Mittelung bereits bei den geschätzten Leistungsdichtespektren von Sprache $\hat{S}_{s_i s_i,k}(n)$ an. Diese werden mit der DAKF gemäß

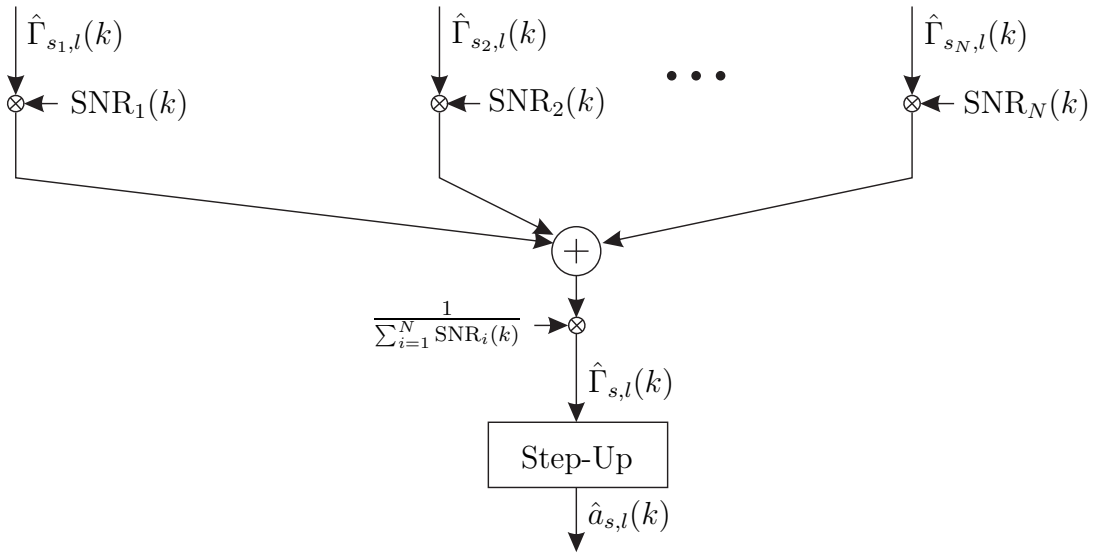

Abbildung 4.5: *Gewichtete Summation der Reflexionskoeffizientensätzen der* Methode 2*. Die Gewichtungsfaktoren bestimmen sich aus dem SNR des jeweiligen Kanals.*

Abbildung 4.3 unter Weglassung des IFFT Blocks berechnet. Die so erhaltenen Periodogramme werden analog zur Bartlett-Methode gemäß Gleichung 4.30 aufaddiert. Der Unterschied besteht allerdings darin, dass zur Mittelung der Periodogramme der Datenblock der Länge L_{AR} nicht in mehrere Segmente der Länge $L_K < L_{\mathrm{AR}}$ zerlegt wird. Stattdessen werden zur Mittelung die Periodogramme $\hat{S}_{s_i s_i,k}(n)$ der einzelnen Kanäle verwendet. Die addierten Periodogramme werden auf ihre Anzahl normiert und mittels IFFT die Autokorrelationsfolge von Sprache $\hat{r}_{ss,k}(l)$ berechnet:

$$\hat{S}_{ss,k}(n) \;=\; \frac{1}{N}\sum_{i=1}^{N}\hat{S}_{s_i s_i,k}(n). \tag{4.54}$$

Im Gegensatz zu den bis jetzt vorgestellten Methoden liegen hier bereits über die Kanäle gemittelte Werte vor, so dass die AR-Koeffizienten direkt mit dem Levinson-Durbin Algorithmus bestimmt werden können. Da sowohl DAKF wie auch Bartlett-Methode stabile AR-Koeffizienten liefern, liegt auch hier stets Stabilität vor. Die *Methode 3* ist in Abbildung 4.6 dargestellt.

Vergleicht man dieses Verfahren mit den bereits vorgestellten Methoden, so sind zwei mögliche Erweiterungen offensichtlich:

- Gewichtung der einzelnen Periodogramme $\hat{S}_{s_i s_i,k}(n)$ mit dem über der Frequenz konstanten Faktor $\mathrm{SNR}_i(k)$ mit anschließender Gewichtung der Summe auf $\sum_{i=1}^{N}\mathrm{SNR}_i(k)$ (analog zum Vorgehen in *Methode 2*).

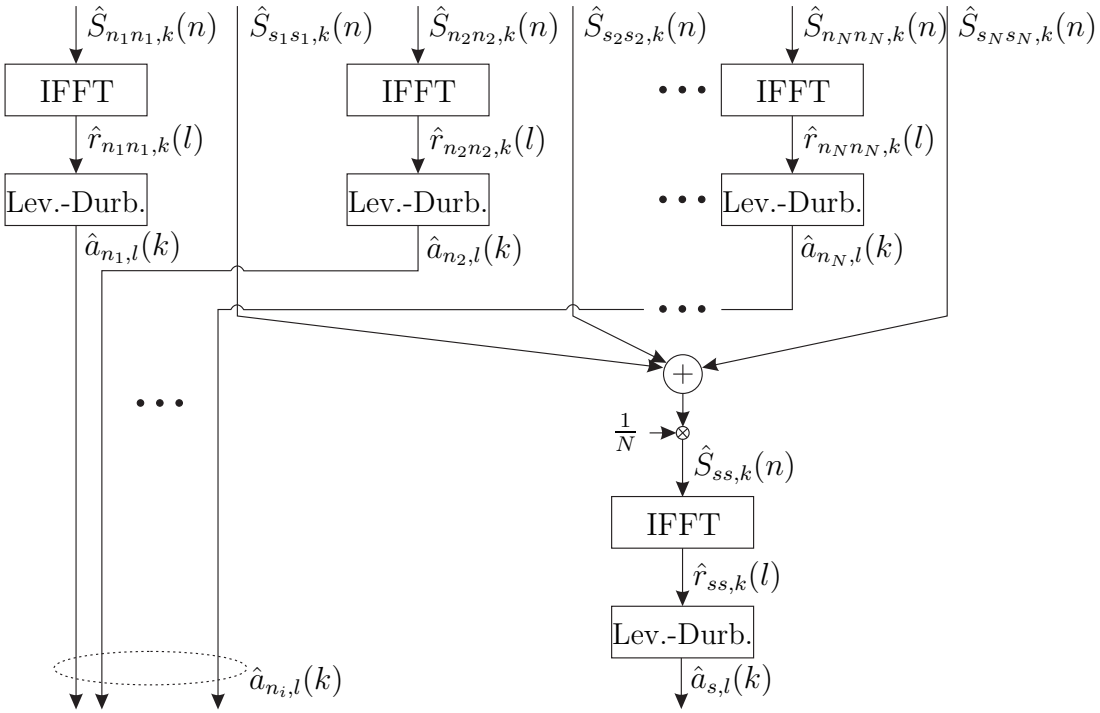

Abbildung 4.6: *Mehrkanalige* Methode 3 *zur Schätzung der AR-Parameter.*

- Verwendung des Welsh-Verfahrens gemäß Gleichung 4.31. Dazu müssen die Autokorrelationsfolgen $\hat{r}_{s_is_i,k}(l)$ vor der Addition ihrer Periodogramme $\hat{S}_{s_is_i,k}(n)$ mit der gewählten Fensterfunktion multipliziert werden.

Letzteres bedeutet allerdings eine zusätzliche Transformation in den Zeitbereich (für die Fensterung) und zurück in den Frequenzbereich (für die Mittelung). Umgehen kann man das Problem durch eine Faltung im Frequenzbereich mit der Transformierten der Fensterfunktion. Jedoch verursachen beide Varianten einen Rechenaufwand, der in keinem Verhältnis zu der erreichbaren Verbesserung der Ergebnisse steht. Daher wird ein dem Welsh-Verfahren ähnliches Vorgehen nicht weiter verfolgt.

Diese Methode verursacht einen deutlich geringeren Aufwand als die beiden zuvor vorgestellten. Zum einen wird bereits bei der N-maligen Berechnung des DAKF-Algorithmus' gespart, da die IFFT am Ende des Signalflussdiagramms aus Abbildung 4.3 weggelassen wurde und nur noch einmal am Ende des Verfahrens berechnet wird. Dies entspricht einer Einsparung von $(N{-}1)\cdot(L_{\mathrm{AR}}{+}p)\ \mathrm{ld}(L_{\mathrm{AR}}{+}p)$ Multiplikationen, wobei $m\,\mathrm{ld}(m)$ den Aufwand für eine FFT/IFFT der Länge m darstellt. Dabei wird allerdings angenommen, dass die Blocklänge m eine Potenz der Zahl Zwei ist. Des Weiteren entfallen $N-1$ Berechnungen des Schur-Algorithmus', weil die Mittelung bereits auf Periodogramm-Ebene durchgeführt wird.

Kombination mit dem Burg-Algorithmus

In [39] wurde die dort vorgestellte DAKF Methode mit dem Burg-Algorithmus kombiniert. Dazu wurde das Eingangssignal $y(k)$ neben der DAKF-Struktur auch dem Burg-Verfahren als Eingangssignal zugeführt. Die resultierenden Reflexionskoeffizienten wurden mit denen der DAKF gemittelt. Diese Mittelung wurde ebenfalls gewichtet durchgeführt, wobei für jedes Teilband ein konstanter Gewichtungsfaktor benutzt wurde.

Es muss beachtet werden, dass der Burg-Algorithmus hier auf das verrauschte Eingangssignal angewendet wird und nicht auf die durch spektrale Subtraktion vom Rauschen getrennte Sprachkomponente. Man benötigt daher keine Sprachpausendetektion und erhält eine Schätzung des verrauschten Sprachsignalspektrums. Allerdings erweist sich hier eine weitere Eigenschaft des Burg-Algorithmus' als nützlich. Dieser bildet eine spektrale Einhüllende mit ausgeprägten Maxima wie bei Sprache deutlich besser nach als eine spektral flache, wie sie bei Rauschen auftritt [37]. Dieser Effekt verstärkt sich insbesondere, wenn die gewählte Ordnung nicht der wirklichen entspricht [26], was in der Praxis meistens der Fall ist. Aufgrund dieser Eigenart erhält man durch Anwendung des Burg-Algorithmus' auf das verrauschte Eingangssignal $y(k)$ bereits eine leichte Geräuschreduktion, solange das SNR nicht zu klein ist [39].

Eine Mittelung dieser Koeffizienten mit denen der DAKF fügt demnach etwas Rauschen in die geschätzten Sprach-AR-Parameter ein. Die Wirkung entspricht der eines Spectral Floors, da damit in der DAKF eventuell entstandene Musical Tones maskiert werden können. Dieses Vorgehen kann alternativ oder in Kombination mit dem bei der spektralen Subtraktion vorgestellten Rauschteppich verwendet werden.

Bei Anwendung im mehrkanaligen Fall gib es zwei mögliche Realisierungen, wobei die erste in Abbildung 4.7 dargestellt ist. Dabei bezeichnen α_{DAKF} und α_{Burg} die Gewichtungsfaktoren der einzelnen Methoden, für deren Summe:

$$\alpha_{\mathrm{DAKF}} + \alpha_{\mathrm{Burg}} \;=\; 1 \tag{4.55}$$

gelten muss, um die Stabilität des AR-Modells zu gewährleisten. Die nach der Schur-Rekursion pro Kanal vorliegenden Sprachreflexionskoeffizienten werden mit denen des Burg-Algorithmus' des entsprechenden Kanals gemittelt, woraus der endgültige Satz des Kanals entsteht:

$$\hat{\Gamma}_{s_i,k}(l) \;=\; \alpha_{\mathrm{DAKF}} \cdot \hat{\Gamma}^{\mathrm{DAKF}}_{s_i,k}(l) + \alpha_{\mathrm{Burg}} \cdot \hat{\Gamma}^{\mathrm{Burg}}_{s_i,k}(l) \tag{4.56}$$

Diese Kombination eignet sich besonders gut für Methoden, bei denen die Mittelung der Koeffizienten über die Reflexionskoeffizienten durchgeführt wird. Das heißt, sie ist nicht geeignet für *Methode 3*, da hier Periodogramme gemittelt

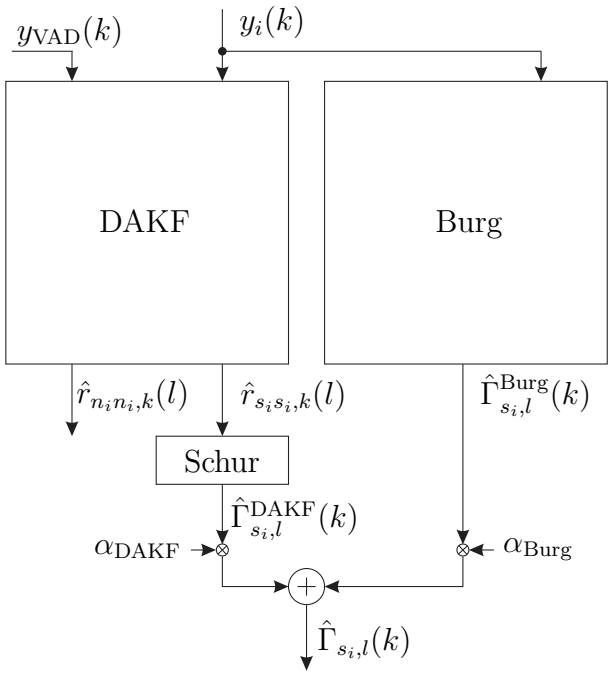

Abbildung 4.7: *Kombination mit dem Burg-Algorithmus innerhalb eines Kanals. Der weitere Signalverlauf der Geräuschkomponente $\hat{r}_{n_i n_i,k}(l)$ ist in diesem Fall nicht von Interesse, weshalb er hier weggelassen wird.*

werden und zusätzliche Umrechnungen zur Integration der Burg-Koeffizienten notwendig wären.

Für die Kombination mit *Methode 3* bietet sich ein anderes Schema an, welches in Abbildung 4.8 dargestellt ist. In deren linken Hälfte ist das Ende des Signalflussdiagramms der *Methode 3* abgebildet. Im Gegensatz zu Abbildung 4.6 werden hier nach der Rücktransformation in den Zeitbereich mittels Schur-Rekursion die Reflexionskoeffizienten $\hat{\Gamma}^{\mathrm{DAKF}}_{s,l}(k)$ bestimmt und nicht direkt die AR-Parameter mittels Levinson-Durbin Algorithmus. Auf der anderen Seite werden pro Kanal die Reflexionskoeffizienten $\hat{\Gamma}^{\mathrm{Burg}}_{s_i,l}(k)$ berechnet und zu einem Satz $\hat{\Gamma}^{\mathrm{Burg}}_{s,l}(k)$ gemittelt.

Diese so erhaltenen Koeffizientensätze werden nun mit den Faktoren α_{DAKF} und α_{Burg} gewichtet und aufaddiert. Diese Kombination kann mit jeder der drei vorgestellten Methoden kombiniert werden.

Unabhängig von der gewählten Variante ist es offensichtlich, dass für die Bestimmung der AR-Parameter des Geräuschs immer die DAKF benötigt wird, da die Burg-Methode in der hier verwendeten Art dies nicht leisten kann.

In dieser Arbeit wurde für *Methode 1* und *Methode 2* die in Abbildung 4.7 darge-

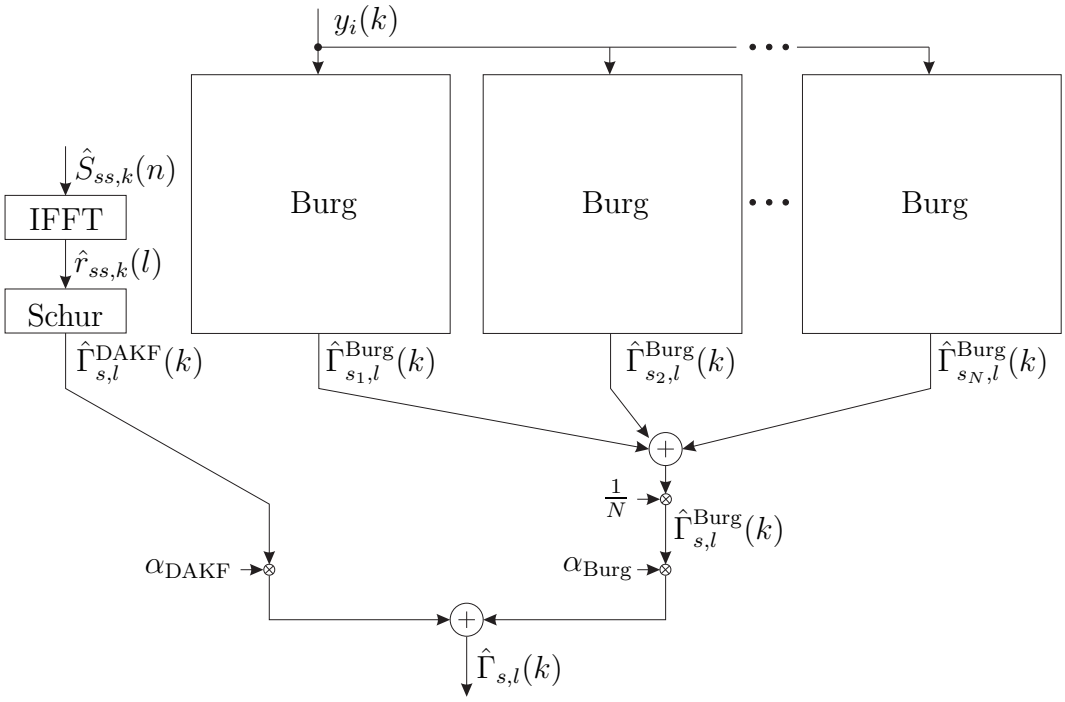

Abbildung 4.8: *Kombination der mittels DAKF bereits gemittelten Koeffizienten mit dem Burg-Algorithmus .*

stellte Möglichkeit der Verknüpfung mit dem Burg-Algorithmus gewählt, während für *Methode 3* der Ansatz aus Abbildung 4.8 verwendet wird. Unabhängig von der benutzten Variante haben sich für den mehrkanaligen Fall und die hier vorliegenden Daten abweichend von [39] die Gewichtungsfaktoren $\alpha_{\mathrm{DAKF}} = 0,8$ und $\alpha_{\mathrm{Burg}} = 0,2$ als günstig erwiesen.

4.4 Schätzung der Raumimpulsantworten und Filterfunktionen

Im letzten Teil dieses Kapitel werden die Möglichkeiten zur Schätzung der in der Messmatrix $\mathbf{C}(k)$ der Messgleichung 3.25 vorkommenden Größen vorgestellt. Dies sind zum einen die Raumimpulsantworten $\mathbf{h}_i(k)$ für $i = 1, \ldots, N$ und zum anderen die Kreuzfilterfunktionen $\mathbf{g}_{i_1 i_2}(k)$ für $i_1 = 1, \ldots, N$, $i_2 = 1, \ldots, N$ und $i_1 \neq i_2$.

4.4.1 Schätzung der Raumimpulsantworten

Von allen in dieser Arbeit vorgestellten Schätzproblemen stellt die Schätzung der Raumimpulsantworten $\mathbf{h}_i(k)$ das schwierigste Problem dar. Dies liegt daran, dass diese *blind* – also ohne über das Mikrofonsignal $y(k)$ hinausgehende Kenntnisse – durchgeführt werden muss. Die für diese Aufgabe heute zur Verfügung stehenden Verfahren, wie zum Beispiel [24, 31], können im Hinblick auf die Verwendung in der Messmatrix $\mathbf{C}(k)$, die für die Matrixinversion bei der Bestimmung der Kalman-Verstärkung verwendet wird, nicht als stabil bezeichnet werden. Darüber hinaus sind die mit ihnen erreichten Ergebnisse bei der Schätzung der Raumimpulsantworten nicht zufriedenstellend, wenn – wie bei dem hier vorgestellten Verfahren – als Eingangsdaten eine verrauschte Sprachquelle angenommen werden muss [31].

Eine vollständige Schätzung der zeitvarianten Raumimpulsantworten scheidet somit aufgrund der verfügbaren Algorithmen aus. Damit ist auch keine vollständige Enthallung möglich. Allerdings gibt es einen weiteren Ansatz, der zumindest eine teilweise Charakterisierung des Raumes erlaubt. In dem in Abschnitt 2.4 beschriebenen Datensatz ist pro Mikrofon auch die mittlere Distanz zum Fahrer vorhanden. In diesem Zusammenhang bezeichnet *mittlere Distanz* diejenige Entfernung, die bei einem mittelgroßen Fahrer auftritt, der den Fahrersitz entsprechend auf eine mittlere Position einstellt. Diese kann benutzt werden, um mit Hilfe der Raumimpulsantworten $\mathbf{h}_i(k)$ einen festen Laufzeitausgleich durchzuführen.

Um einen solchen Laufzeitausgleich herzustellen, müssen die Eingangssignale um nicht-ganzzahlige Vielfache des Abtastintervalls $T = 1/f_s$ verzögert werden. Die Laufzeit ΔT_i des Signals zu den einzelnen Mikrofonen berechnet sich über die mittlere Distanz $\bar{d}_i$ sowie die Schallgeschwindigkeit $c_{\text{Schall}} = 330$ m/s zu:

$$\Delta T_i = \frac{\bar{d}_i}{c} \quad \text{für} \quad i = 1, \ldots, N. \tag{4.57}$$

Ziel eines Laufzeitausgleichs ist es nun, alle Eingangssignale, bis auf das Signal mit der längsten Laufzeit, so zu verzögern, dass nach der zusätzlichen Verzögerung alle Laufzeiten identisch der maximalen Laufzeit $\Delta T_{\max} = \max\{\Delta T_i\}$ für $i = 1, \ldots, N$ sind.

Aufgrund der in Kapitel 3 gewählten Struktur müssen diese Laufzeit einfügenden Filter, im Folgenden als Laufzeitfilter bezeichnet, als Transversalfilter (FIR) ausgelegt sein. Ein solcher Filtertyp wird im Englischen als *Fractional Delay Filter* bezeichnet. Zu dessen Berechnung gibt es verschiedene Möglichkeiten, die beispielsweise in [29, 50, 51] gefunden werden können. Eine Untersuchung dieser Filterdesignansätze für die Implementierung eines Griffith-Jim-Beamformer im Vollband findet sich zudem in [57]. Von den genannten Ansätzen ist besonders die *Farrow*-Struktur [29] interessant, da sie eine Veränderung der Verzögerung

ohne Neuberechnung der Laufzeitfilter erlaubt. Moderne Oberklasse-Fahrzeuge speichern die Position des Fahrersitzes für die nächste Benutzung. Daher ist es denkbar, diese Information in zukünftigen Verfahren für die Anpassung der Filter – auch während der Fahrt – zu verwenden.

Oben genannte Verfahren liefern die Impulsantworten der gesuchten Laufzeitfilter im Vollband. Für die in dieser Arbeit vorgeschlagene Filterbankstruktur müssen diese ins Teilband transformiert werden. Dabei entstehen, selbst wenn von einem kausalen Filter ausgegangen wird, nicht-kausale Anteile [44], welche allerdings aufgrund der Struktur der Messmatrix nicht berücksichtigt werden können, da diese nur kausale Anteile vorsieht. Das gleiche gilt für die im nächsten Abschnitt vorgestellten Kreuzfilterfunktionen.

4.4.2 Schätzung der Kreuzfilter

Die Impulsantworten der Kreuzfilter $\mathbf{g}_{i_1 i_2}(k)$ stellen den Zusammenhang zwischen den verschiedenen Rauschquellen dar. Dieser Zusammenhang ist nur vorhanden, wenn die Quellen zueinander korrelierte Anteile aufweisen (siehe Abschnitt 3.1). Das im Folgenden vorgestellte Schätzverfahren wird am Beispiel des Kreuzfilters $\mathbf{g}_{12}(k)$ hergeleitet, um die umständliche Notation der allgemeinen Mikrofonindizes i_1 und i_2 zu vermeiden.

Für den hier gewählten Fall soll $\mathbf{g}_{12}(k)$ so bestimmt werden, dass die Faltung $\sum_{l=0}^{q-1} g^*_{12,l}(k) n_1(k-l)$ den zu $n_1(k)$ korrelierten Anteil in $n_2(k)$ ergibt, wobei q die Ordnung des Geräuschmodells bezeichnet. Vorausgesetzt die Signale $n_1(k)$ und $n_2(k)$ stehen zur Verfügung, kann die Impulsantwort $\mathbf{g}_{12}(k)$ mit einem adaptiven Filter eingestellt werden. Aufgrund seiner Robustheit wird zum Einstellen dieses Filters der NLMS (Normalized Least Mean Square) Algorithmus verwendet [19, 22, 43]. Die benutzte Struktur ist in Abbildung 4.9 für die Kombination von $\mathbf{g}_{12}(k)$ und $\mathbf{g}_{21}(k)$ dargestellt.

Doch zunächst stellt sich das Problem, dass die dafür notwendigen Geräuschsignale $n_1(k)$ und $n_2(k)$ nicht direkt gemessen werden können, da nur die entsprechenden Mikrofonsignale $y_1(k)$ und $y_2(k)$ zur Verfügung stehen. Liegt allerdings eine Sprachpause vor, so gilt:

$$n_1(k) = y_1(k) \quad \text{und} \quad n_2(k) = y_2(k) \quad \text{für} \quad y_{\mathrm{VAD}}(k) = 0, \tag{4.58}$$

und die Geräuschsignale sind direkt messbar. Für die benötigte Sprachpausendetektion kann die in der DAKF bereits vorhandene verwendet werden. Während einer Sprachphase wird die zuletzt eingestellte Kreuzfilterfunktion konstant gehalten, während einer Sprachpause wird sie entsprechend adaptiert:

$$\mathbf{g}_{12}(k) \;=\; \begin{cases} \mathbf{g}_{12}(k-1) & \text{für} \quad y_{\mathrm{VAD}}(k) = 1 \\ \mathbf{g}_{12}(k-1) + \alpha_{\mathrm{NLMS}} \frac{\mathbf{y_1}(k) e^*_{12}(k)}{\|\mathbf{y_1}(k)\| + \Delta} & \text{für} \quad y_{\mathrm{VAD}}(k) = 0 \end{cases}, \tag{4.59}$$

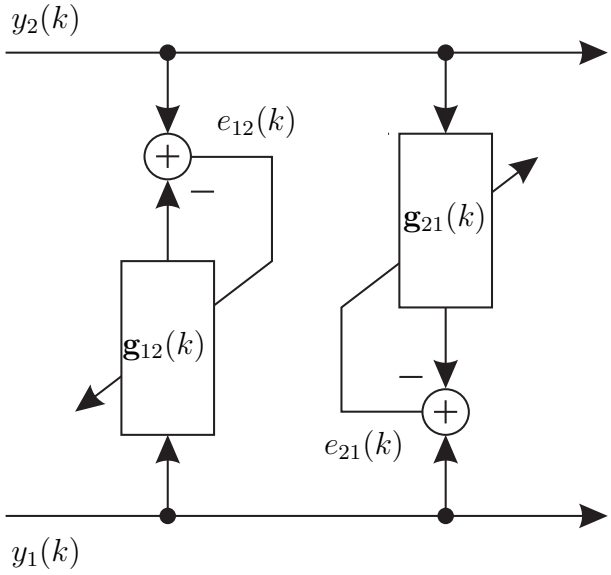

Abbildung 4.9: *Struktur für die Schätzung der Kreuzfilterfunktionen mittels adaptiver Filter.*

wobei $l = 0, \ldots, q-1$ gilt. Der Adaptionsfehler ergibt sich zu:

$$e_{12}(k) \quad = \quad y_2(k) - \mathbf{g}_{12}^{\mathrm{H}}(k)\mathbf{y}_1(k). \tag{4.60}$$

Um die Stabilität der Adaption sicherzustellen, müssen zwei Punkte beachtet werden, da die hier verwendeten Filterlängen sehr klein sind[7] und nur ein Teil von $n_2(k)$ – nämlich der zu $n_1(k)$ korrelierte – adaptiert werden soll:

- Die Schrittweite α_{NLMS} muss klein genug gewählt werden, um ausreichend robust gegenüber Fehladaptionen zu sein. Werte im Bereich 0.01 bis 0.1 haben sich hierfür als günstig erwiesen.
- Aufgrund der kurzen Filterlänge muss der Nenner in Gleichung 4.59 regularisiert (das heißt $\Delta > 0$) werden [18].

Da $y_1(k)$ und $y_2(k)$ nicht laufzeitausgeglichen sind, weisen die Kreuzfilterfunktionen nicht-kausale Anteile auf. Aufgrund der Filterbankimplementierung und der dadurch vorgenommenen Unterabtastung ist der Laufzeitunterschied relativ zum Abtastintervall sehr klein, weshalb ein vorgeschalteter Laufzeitausgleich an dieser Stelle vernachlässigt werden kann.

[7] In Kapitel 5 wird maximal $q = 4$ verwendet

Kapitel 5

Implementierung des Gesamtsystems

Nachdem in Kapitel 3 das verwendete Kalman-Filter hergeleitet wurde und die Schätzung der zum Betrieb notwendigen Parameter im letzten Kapitel vorgestellt wurde, werden in diesem Kapitel diejenigen Themen diskutiert, welche die Implementierung betreffen. Dies geschieht in zwei Schritten: Im ersten Teil wird die Signalverarbeitung mittels Polyphasen-Filterbänken erläutert. Diese stellt einen Zwitter aus Zeit- und Frequenzbereichsverarbeitung dar, was im nächsten Abschnitt näher erläutert wird. Der zweite Teil des Kapitels beschäftigt sich mit Parametrierung der zeitinvarianten Größen, insbesondere mit den Ordnungen der Sprach- und Geräuschmodelle.

5.1 Verarbeitung mittels Polyphasen-Filterbank

Dieser Abschnitt beschäftigt sich mit der verwendeten Filterbankstruktur. Dazu wird zunächst ein kurzer Überblick über die grundlegenden Verarbeitungsmethoden auf dem Gebiet der Signalverarbeitung gegeben. Anschließend wird die Verarbeitung mittels Filterbänken insbesondere mittels Polyphasen-Filterbänken beschrieben.

5.1.1 Gebräuchliche Signalverarbeitungsstrukturen

Bei der Auslegung von Signalverarbeitungssystemen können für die Implementierung prinzipiell drei Hauptstrukturen unterschieden werden [18]. Man teilt ein in Verfahren mit Verarbeitung im:

- Zeitbereich,
- Frequenzbereich
- und Teilbandbereich.

Für die Bewertung der verschiedenen Strukturen werden vier Kriterien verwendet, wobei kein Verfahren in jeweils allen Kategorien optimal sein kann. Die Bewertungskriterien sind:

1. Signalverzögerung,
2. Rechenaufwand,
3. Zeitauflösung
4. und Frequenzauflösung.

Zeitbereichsverarbeitung

Bei der Verarbeitung im Zeitbereich, im Englischen als *Fullband Processing* bezeichnet, kann die Parametrierung des Algorithmus' für jeden Zeitschritt bzw. jeden Abtastwert verändert werden. Dadurch wird eine hohe Zeitauflösung erreicht. Auf der anderen Seite wirkt sich eine stetige Veränderung der Zeitbereichsparameter sozusagen auf alle Frequenzen aus, woraus eine geringe Frequenzauflösung resultiert. Da mit jedem neuen Abtastwert der jeweilige Algorithmus ausgeführt wird, ist der Rechenaufwand groß. Andererseits müssen keine Eingangswerte für eine blockweise Verarbeitung gesammelt werden, wodurch zum einen die Signalverzögerung gering ausfällt und zum anderen kein Speicher für eine Blockbildung vorgehalten werden muss.

Frequenzbereichsverarbeitung

Im Gegensatz dazu wird für die Verarbeitung im Frequenzbereich (englisch: *Block Processing*) das Eingangssignal in gleich große Abschnitte zerlegt, die anschließend jeweils als Block verarbeitet werden. Dadurch entstehen eine zur Blocklänge korrespondierende Signalverzögerung und ein erhöhter Speicherbedarf. Die zeitliche Auflösung ist entsprechend verschlechtert. Auf der anderen Seite wird durch diese Blockverarbeitung der Rechenaufwand stark reduziert und die Manipulation der in den Frequenzbereich transformierten Blöcke ergibt eine hohe Frequenzauflösung.

Teilbandverarbeitung

Die Verarbeitung im Teilbandbereich (englisch: *Subband Processing*) kann in der Mitte zwischen Zeit- und Frequenzbereichsverarbeitung eingeordnet werden, da sie positive Eigenschaften von beiden Strukturen aufweist. Zum einen ist es möglich, sowohl eine akzeptable Frequenz- als auch Zeitauflösung zu erreichen. Darüber hinaus kann die Aufwandseinsparung der Blockverarbeitung gegenüber der Vollbandverarbeitung zum Teil übernommen werden, während die verursachte Signalverzögerung dagegen weniger stark ausgeprägt ist. Zudem ermöglicht die Teilbandstruktur die Anwendung von Zeitbereichs- wie auch Frequenzbereichsalgorithmen und weist dadurch aus Sicht des Systemdesigners die meisten Freiheitsgrade auf. Diese Struktur wird im Folgenden detailiert erläutert.

5.1.2 Struktur der Teilbandverarbeitung

In Abbildung 5.1 ist das Prinzip der Teilbandverarbeitung dargestellt. Das breitbandige Eingangssignal $y(k)$ wird in der sogenannten *Analysefilterbank* in M schmalbandige, im Spektralbereich aneinander angrenzende und teilweise überlappende Frequenzbandsignale $Y_\mu(k')$ mit $\mu = 0, \ldots, M-1$ aufgespalten. Im Teilbandbereich können diese Signale anschließend einer Unterabtastung um den Faktor r unterzogen werden, was durch die Verwendung des Zeitindex' k' im Teilband dargestellt wird. Die entstehenden Teilbandsignale sind mathematisch formal immer noch Zeitbereichssignale, wobei jedes Band einen spektralen Ausschnitt des ursprünglichen Vollbandsignals enthält. Für einen gegebenen Zeitpunkt $k' = k'_0$ kann die Schar der Signale $Y_\mu(k'_0)$ dagegen als Frequenzbereichssignal interpretiert werden, weshalb sie in Großbuchstaben sowohl mit Zeitindex k' als auch mit Teilbandindex μ notiert werden.

Die Signale am Ausgang der Analysefilterbank werden dem eigentlichen Verfahren – im Beispiel aus Abbildung 5.1 einer Geräuschreduktion – zugeführt. Die Verarbeitung wird im Teilbandbereich durchgeführt. Dabei ermöglicht die zum Vollband relative Schmalbandigkeit der einzelnen Teilbandsignale, dass gegenüber einer Vollbandimplementierung beispielsweise bei einem adaptiven Filter Filterlänge eingespart und bei einem linearen Prädiktor eine kleinere Ordnung gewählt werden kann. Bei Verfahren, deren Rechenaufwand schneller als linear mit der Filterlänge oder Ordnung wächst (z.B. Levinson-Durbin-Algorithmus), lassen sich so deutliche Aufwandseinsparungen erreichen.

Ähnliches gilt für das in Kapitel 3 hergeleitete MISO Kalman-Filter: Bei einer Vollbandimplementierung muss für eine Modellierung aller Pitchkomponenten die Sprachmodellordnung p im Bereich von 60 bis 80 gewählt werden. Dies kann anhand Abbildung 2.4 überprüft werden. Der dort abgebildete stimmhafte Si-

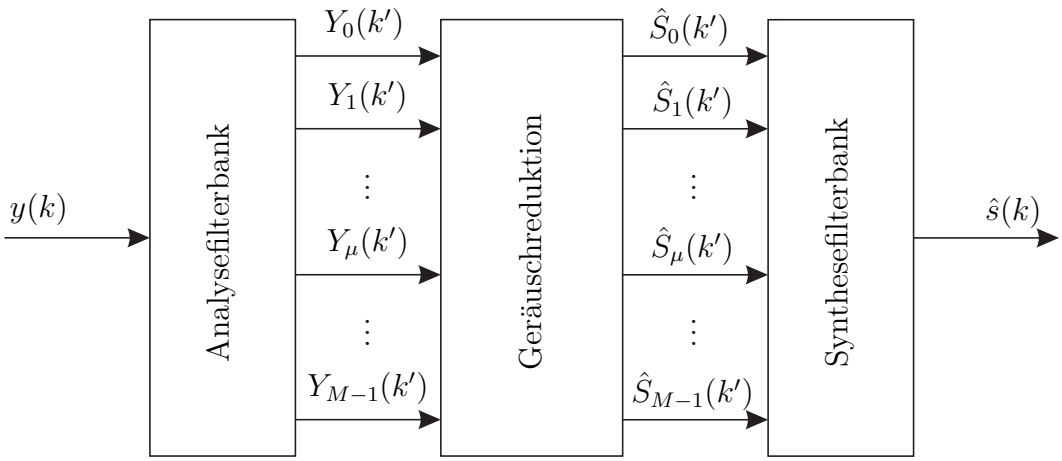

Abbildung 5.1: *Aufbau eines Geräuschreduktionsverfahrens eingebettet in Analyse- und Synthesefilterbank. Eine mögliche reduzierte Abtastrate im Teilbandbereich wird mit dem Zeitindex k' angedeutet.*

gnalausschnitt weist etwas mehr als 30 Pitchmaxima auf. Für deren Modellierung müsste, da reellwertige Signale vorliegen, die doppelte Ordnung benutzt werden. Die Verwendung einer derart hohen Sprachmodellordnung macht das Kalman-Filter sehr unhandlich, da der numerische Aufwand kubisch mit der Ordnung wächst (siehe Abschnitt 3.5), und für die Anzahl an Rechenoperationen pro Abtastzyklus aufgrund der hier verwendeten Abtastrate von $f_s = 8000$ Hz nur $1/f_s = 125$ μs zur Verfügung stehen. Ein Lösungsmöglichkeit besteht darin, nur noch die Formantstruktur des Sprachsignals zu modellieren, was zu einem Informations- und somit Qualitätsverlust führt. Dieser Vollbandansatz mit Formantmodellierung wird beispielsweise in [16, 30, 33] verwendet. Im Gegensatz dazu ermöglicht die in dieser Arbeit benutzte Teilbandverarbeitung, dass die Sprachmodellordnung pro Frequenzband zwar klein bleibt ($p \leq 6$), auf das Vollband hochgerechnet aber alle Pitchkomponenten modelliert werden können.

Die im Teilband berechneten Schätzwerte $\hat{S}_\mu(k')$ für die Sprachkomponente müssen nun wieder zu einem Vollbandsignal zusammengesetzt werden, was in der sogenannten *Synthesefilterbank* geschieht. An deren Ausgang liegt der gesuchte Schätzwert $\hat{s}(k)$ der Sprachkomponente im Zeitbereich an.

Im Folgenden werden die mathematischen Zusammenhänge von Analyse- und Synthesefilterbank beschrieben. Dabei wird ausschließlich auf die im vorgeschlagenen Verfahren verwendete *Polyphasen-Filterbank* eingegangen. Andere Filterbankstrukturen wie beispielsweise QMF-Filterbänke[1] oder Filterbänke mit un-

[1] QMF steht für *Quadrature-Mirror-Filters* und bezieht sich auf die paarweise und symmetrische Anordnung von Hoch- und Tiefpassfilter. Kaskadiert angeordnet können so Filterbänke verschiedener Bandanzahl und Teilband-Bandbreite erzeugt werden.

gleichmäßiger Frequenzaufteilung (im Englischen als *Non-Uniform Filterbanks* bezeichnet) finden sich unter anderem in [54, 55].

5.1.3 Polyphasen-Filterbänke

Die Polyphasen-Filterbank gehört zu der Klasse der DFT-modulierten Filterbänke und stellt eine Verallgemeinerung der Transformation mittels DFT und vorangegangener Fensterung dar [3, 11, 14, 52]. Die nachfolgende Herleitung orientiert sich dabei an [39, 44].

Zunächst wird das Eingangssignal $y(k)$ der Bandbreite B durch ein Tief-, mehrere Band- und ein Hochpassfilter in M Frequenzbänder zerlegt. Dabei wird angenommen, dass M gerade ist[2]. Die Signale $Y_\mu(k)$ am Ausgang dieser Filter nehmen bei ideal frequenzselektiver Filterung (d.h. Rechteckfilterung) nur noch den M-ten Teil der ursprünglichen Bandbreite B ein, während Artefakte der benachbarten Bänder vollständig unterdrückt werden (ideale Sperrdämpfung). Entsprechend kann in einem zweiten Schritt die Abtastrate um den Faktor $r \leq M$ reduziert werden. Dadurch wird erreicht, dass der durch die parallele Verarbeitung von M Bändern entstehende Mehraufwand teilweise oder ganz kompensiert wird.

Die maximal mögliche Abtastratenreduzierung von $r = M$ wird hierbei als *kritische Unterabtastung* bezeichnet. In diesem Fall kompensieren sich notwendiger Mehraufwand und erreichbare Einsparung exakt. Damit bei kritischer Unterabtastung keine Überfaltungseffekte (englisch: *Aliasing*) zwischen benachbarten Frequenzbändern entstehen, müssen die vorausgegangenen Filterungen – wie bereits oben erwähnt – ideal frequenzselektiv sein. Dies ist in der Realität unmöglich, da reale Impulsantworten stets kausal und zeitbegrenzt sein müssen, was eine Rechteckfilterung ausschließt. In der Anwendung wird daher die Unterabtastrate meistens $r < M$ gewählt. Dabei muss stets ein Kompromiss zwischen wenig Aliasing (r möglichst klein) und großer Rechenaufwandseinsparung (r möglichst groß, d.h. nahe bei M) gefunden werden.

Die Besonderheit der Polyphasen-Filterbank ist, dass Tief-, Band- und Hochpassfilter durch Frequenzverschiebung von ein und demselben Grundfilter, dem sogenannten *Prototyp-Tiefpassfilter* h_k, abgeleitet werden. Für die Filter $h_{\mu,k}$ der Analysefilterbank gilt:

$$h_{\mu,k} \quad = \quad h_k e^{j\frac{2\pi}{M}\mu k}. \tag{5.1}$$

Die einseitigen Bandbreiten des Tiefpassfilters ($\mu = 0$) und des Hochpassfilters

[2]Die in diesem Abschnitt hergeleiteten Bedingungen und Gleichungen unterscheiden sich für gerade und ungerade M, daher ist es zweckmäßig, sich zur Vereinfachung der Darstellung im Folgenden auf eines festzulegen.

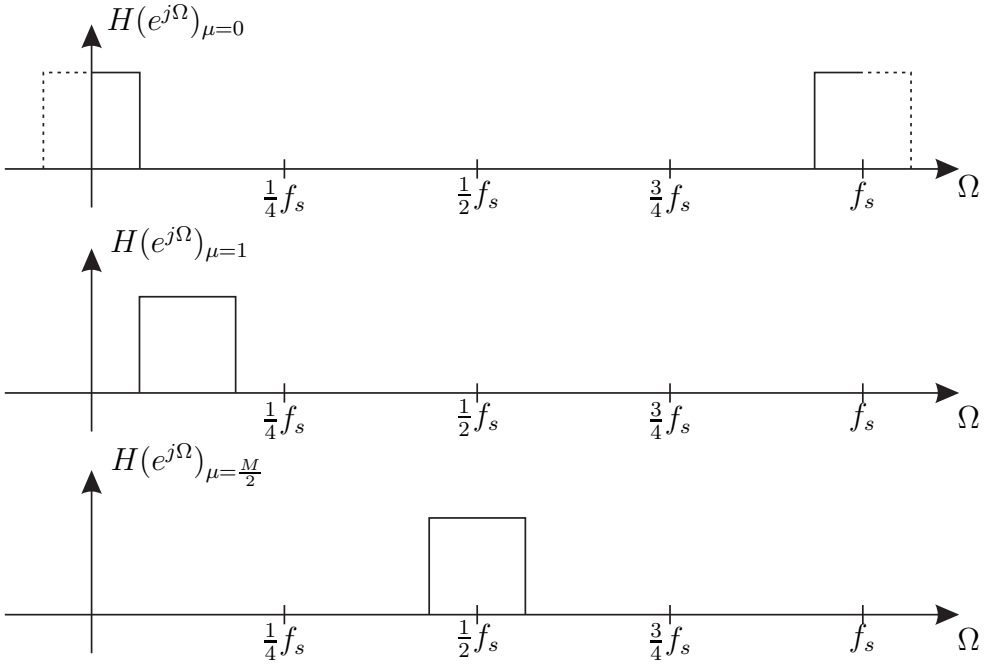

Abbildung 5.2: *Frequenzgang von Prototyptiefpassfilter ($\mu = 0$), erstem Bandpassfilter ($\mu = 1$) und Hochpassfilter ($\mu = \frac{M}{2}$) bei kritischer Unterabtastung und ideal frequenzselektiver Filterung.*

($\mu = \frac{M}{2}$) ergeben sich zu:

$$B_{\mathrm{TP}} = B_{\mathrm{HP}} \quad = \quad \frac{f_s}{2M}, \tag{5.2}$$

die Bandbreite der Bandpassfilter ($\mu = 1, \ldots, \frac{M}{2} - 1$) beträgt entsprechend:

$$B_{\mathrm{BP}} \quad = \quad \frac{f_s}{M}. \tag{5.3}$$

Der Zusammenhang zwischen Prototyptiefpassfilter und den durch Verschiebung daraus entstehenden Band- und Hochpassfiltern ist in Abbildung 5.2 schematisch für kritische Unterabtastung und Rechteckfilterung dargestellt.

Analysefilterbank

Die Filterung für das μ-te Teilband kann als Faltung des Eingangssignals $y(k)$ mit dem frequenzverschobenen Prototyptiefpassfilter h_k aus Gleichung 5.1 geschrie-

ben werden:

$$\begin{aligned} Y_\mu(k) &= \sum_{\kappa=-\infty}^{\infty} h_{\mu,\kappa} y(k-\kappa) \\ &= \sum_{\kappa=-\infty}^{\infty} h_\kappa e^{j\frac{2\pi}{M}\mu\kappa} y(k-\kappa), \qquad \mu = 0,\ldots,M-1. \end{aligned} \tag{5.4}$$

Wie zuvor beschrieben, können die einzelnen Teilbandsignale um den Faktor r unterabgetastet (k'=rk) werden:

$$Y_\mu(k') = \sum_{\kappa=-\infty}^{\infty} h_\kappa e^{j\frac{2\pi}{M}\mu\kappa} y(rk-\kappa), \qquad \mu = 0,\ldots,M-1. \tag{5.5}$$

Fordert man zusätzlich, dass die Impulsantwort h_k kausal und zeitbegrenzt ist, können die Grenzen der Summe aus Gleichung 5.5 entsprechend eingeschränkt werden. Für die Länge L_{PLP} des Prototyptiefpassfilters wird dabei angenommen, dass sie ein ganzzahliges Vielfaches der Bandanzahl M ist: $L_{\mathrm{PLP}} = K \cdot M$. Andernfalls kann ohne Einschränkung der Allgemeinheit die Länge von h_k durch Anhängen von Nullen angepasst werden. Man erhält:

$$Y_\mu(k') = \sum_{\kappa=0}^{L_{\mathrm{PLP}}-1} h_\kappa e^{j\frac{2\pi}{M}\mu\kappa} y(rk-\kappa), \qquad \mu = 0,\ldots,M-1. \tag{5.6}$$

Durch die Substitution

$$\kappa = \lambda M + \nu \tag{5.7}$$

mit $\lambda = 0,\ldots,K-1$ und $\nu = 0,\ldots,M-1$ kann Gleichung 5.5 in die effiziente Struktur der Polyphasenfilterbank überführt werden:

$$\begin{aligned} Y_\mu(k') &= \sum_{\nu=0}^{M-1}\sum_{\lambda=0}^{K-1} h_{\lambda M+\nu} \underbrace{e^{j\frac{2\pi}{M}\mu\lambda M}}_{=1} e^{j\frac{2\pi}{M}\mu\nu} y(rk-\lambda M-\nu) \\ &= \sum_{\nu=0}^{M-1} \underbrace{\sum_{\lambda=0}^{K-1} y(rk-\lambda M-\nu) h_{\lambda M+\nu}}_{y_\nu(rk)} e^{j\frac{2\pi}{M}\mu\nu} \\ &= \sum_{\nu=0}^{M-1} y_\nu(rk) e^{j\frac{2\pi}{M}\mu\nu} = M \cdot \mathrm{IDFT}\{y_\nu(rk)\} \end{aligned} \tag{5.8}$$

Die Struktur der Analysefilterbank aus Gleichung 5.8 ist als Blockschaltbild in Abbildung 5.3 dargestellt. Auffallend dabei ist insbesondere, dass die Länge

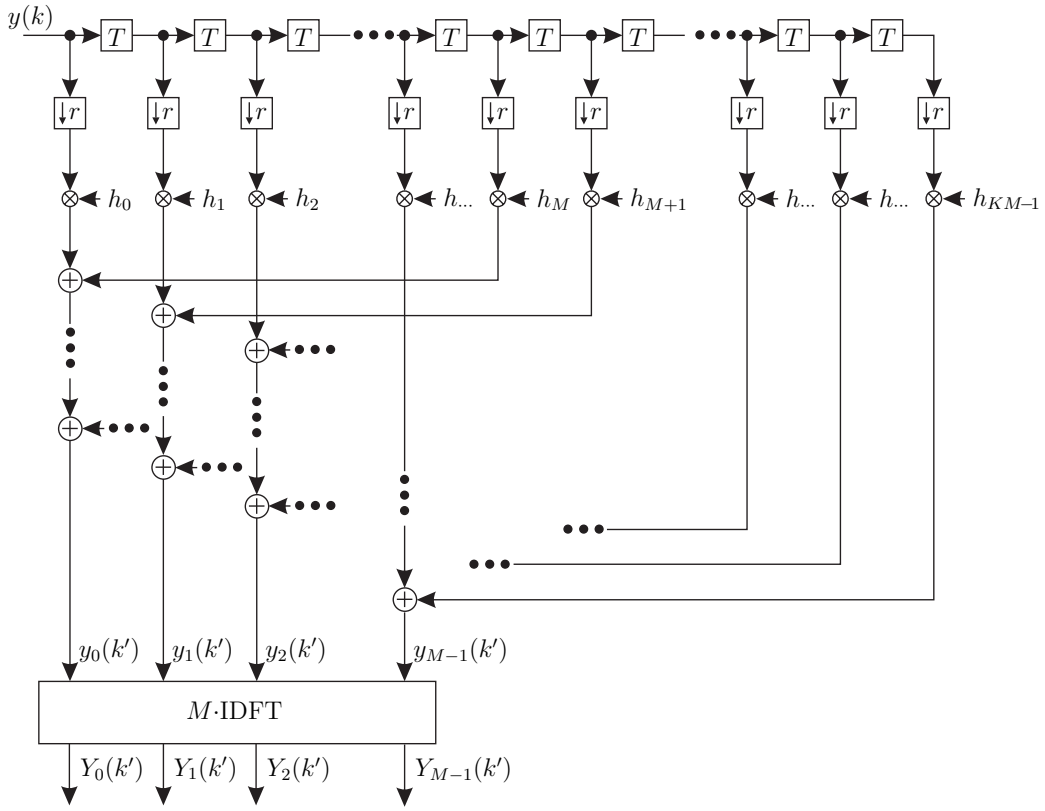

Abbildung 5.3: *Analysefilterbank in Polyphasenstruktur.*

L_{PLP} des Prototyptiefpass' größer als die Bandanzahl M ist. Dadurch kann eine höhere Flankensteilheit (Frequenzselektivität) und Nebenmaximaunterdrückung (Sperrdämpfung) als mit Filtern der Länge M erreicht werden. Insofern stellt die Polyphasenfilterbank eine Verallgemeinerung der Kurzzeitsprektralanalyse mittels DFT und vorangegangener Fensterung (*Fenster-DFT*) dar bzw. kann die Fenster-DFT als Sonderfall der Polyphasenfilterbank interpretiert werden [52, 53].

Synthesefilterbank

Die in M Bänder aufgespaltenen und unterabgetasteten Teilbandsignale $Y_\mu(k')$ mit $\mu = 0, \ldots, M-1$ werden nun im Teilbandbereich der eigentlichen Geräuschreduktion unterzogen. Diese liefert für jedes Teilband einen Schätzwert $\hat{S}_\mu(k')$ der jeweiligen Sprachkomponente $S_\mu(k')$, so dass wie bereits in Abbildung 5.1 dargestellt gilt:

$$\hat{S}_\mu(k') = \mathrm{NR}\left\{Y(k')_\mu\right\} \quad \text{mit} \quad Y_\mu(k') = S_\mu(k') + N_\mu(k'), \tag{5.9}$$

wobei der Operator NR{ } stellvertretend für alle zur Geräuschreduktion durchgeführten Maßnahmen steht und $\mu = 0, \ldots, M-1$ gilt.

Aufgabe der Synthesefilterbank ist es nun, die nach der Teilbandverarbeitung vorliegenden Signale der einzelnen Bänder $\hat{S}_\mu(k')$ wieder zu einem Vollbandsignal $\hat{s}(k)$ zu kombinieren. Dazu werden im Prinzip die Schritte der Analysefilterbank in umgekehrter Reihenfolge durchgeführt, weshalb auf eine ausführliche Herleitung verzichtet wird. Detaillierte Beschreibungen finden sich beispielsweise in [44, 55].

Das unterabgetastete Teilbandsignal $\hat{S}_\mu(k')$ wird zunächst um den Faktor r überabgetastet ($k = \frac{k'}{r}$). Dadurch wird dessen Spektrum auf den r-ten Teil seiner ursprünglichen Breite gestaucht. Zusätzlich entstehen $r-1$ Spiegelspektren, die durch ein sogenanntes *Anti-Imaging Filter* $g_{\mu,k}$ entfernt werden müssen. Danach können die so bearbeiteten Bandsignale zum Vollbandsignal aufaddiert werden:

$$\hat{s}(k) \;=\; \sum_{\mu=0}^{M-1} \sum_{\kappa=-\infty}^{\infty} \hat{S}_\mu(\frac{k'-\kappa}{r}) g_{\mu,\kappa}. \tag{5.10}$$

Dieses bandabhängige Filter $g_{\mu,k}$ kann genau wie bei der Analyse als kausales, frequenzverschobenes Prototyptiefpassfilter g_k der Länge $L_{\mathrm{PLP}} = K \cdot M$ realisiert werden:

$$\hat{s}(k) \;=\; \sum_{\mu=0}^{M-1} \sum_{\kappa=0}^{L_{\mathrm{PLP}}-1} \hat{S}_\mu(\frac{k'-\kappa}{r}) g_\kappa e^{j\frac{2\pi}{M}\mu\kappa}. \tag{5.11}$$

Mit der Substitution $k' = \lambda r + \nu$ erhält man wiederum eine effiziente Implementierungsstruktur, die in Abbildung 5.4 dargestellt ist.

Der Hauptvorteil dieser effizienten Strukturen für Analyse und Synthese besteht darin, dass die IDFTs, die mittels der schnellen inversen Fouriertransformation (IFFT) realisiert werden, jeweils nur alle r Takte berechnet werden müssen, was weitere Rechenleistung einspart.

Zudem übertragen sich einige Eigenschaften der DFT auf die Polyphasenfilterbank. So kann bei reellen Eingangssignalen aufgrund der Symmetrie der Summe innerhalb der IDFT die obere Hälfte der Teilbänder weggelassen werden, da

$$Y_\mu(k') = Y^*_{M-\mu}(k') \quad \text{für} \quad \mu = 0, \cdots, \frac{M}{2} \tag{5.12}$$

gilt. Nach der Analyse brauchen demnach nur die Bänder $\mu = 0, \cdots, \frac{M}{2}$ weiter im Teilbandbereich verarbeitet werden. Vor der Synthese werden die fehlenden Bänder aus den konjugiert-komplexen, durch Spiegelung der in der Reihenfolge umgekehrten Signalen der unteren Bänder berechnet. Weiterhin sind in diesem Fall die Teilbandsignale für $\mu = 0$ und $\mu = \frac{M}{2}$ rein reell.

In der gewählten Herleitung wird für Analyse- und Synthesefilterbank die IDFT benutzt. Alternativ kann für die Analysestufe auch eine auf der DFT basierende

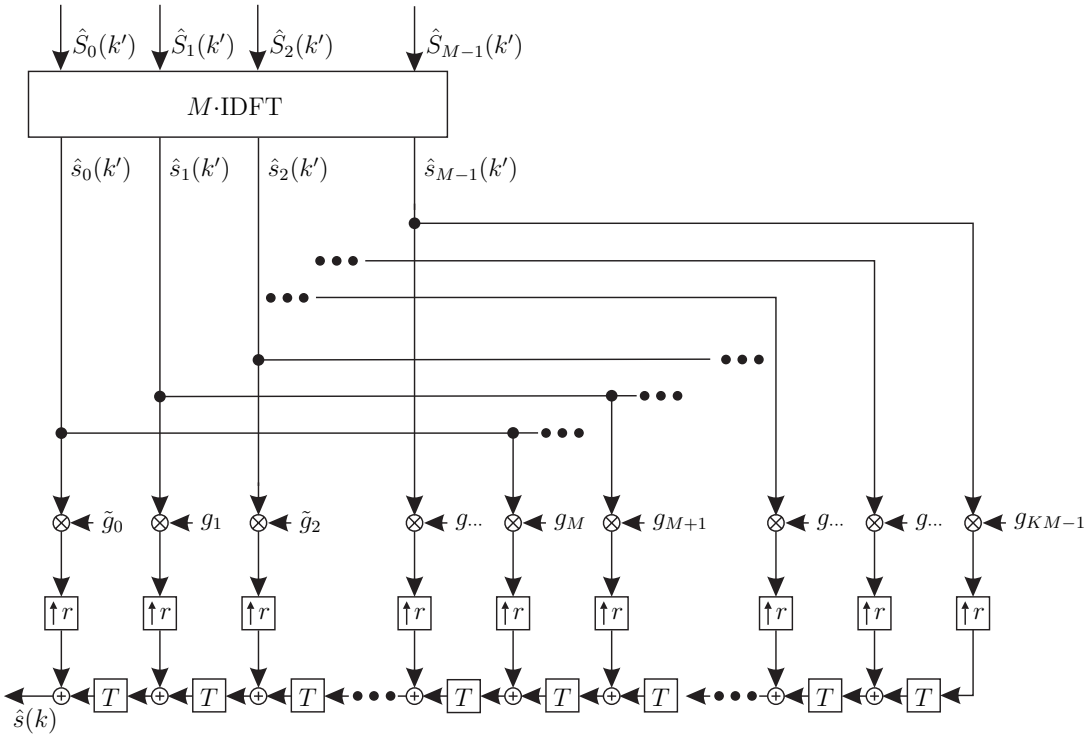

Abbildung 5.4: *Synthesefilterbank in Polyphasenstruktur.*

Struktur hergeleitet werden. Dazu wird die Substitution aus Gleichung 5.7 durch $\kappa = \lambda M - \nu$ ersetzt. Man erhält in diesem Fall nach analoger Rechnung die gesuchte auf diskreter Fouriertransformation basierende Struktur mit einer anderen Reihenfolge der DFT-Eingänge [39].

5.1.4 Betrachtung im Spektralbereich

Um die Auswirkung der Filterung mit dem Prototyptiefpassfilter sowie der Unterabtastung um den Faktor r beschreiben zu können, bietet sich die Darstellung im Frequenzbereich an.

Mit Hilfe der endlichen geometrischen Reihe kann Gleichung 5.6 im Spektralbereich notiert werden:

$$\begin{aligned} Y_\mu(e^{j\Omega}) &= \sum_{k'=-\infty}^{\infty} Y_\mu(k') e^{-j\Omega k'} \\ &= \frac{1}{r} \sum_{\xi=0}^{r-1} Y_\mu\left(e^{j\left(\frac{\Omega}{r} - \frac{2\pi}{r}\xi\right)}\right) H_\mu\left(e^{j\left(\frac{\Omega}{r} - \frac{2\pi}{r}\xi - \frac{2\pi}{M}\mu\right)}\right). \end{aligned} \tag{5.13}$$

Analog hierzu kann Gleichung 5.11 ebenfalls im Frequenzbereich geschrieben wer-

den:

$$\begin{aligned}\hat{S}_\mu(e^{j\Omega}) &= \sum_{k=-\infty}^{\infty} \hat{s}(k)e^{-j\Omega k} \\ &= \sum_{\mu=0}^{M-1} \hat{S}_\mu\left(e^{j\Omega r}\right) G_\mu\left(e^{j\left(\Omega-\frac{2\pi}{M}\mu\right)}\right). \end{aligned} \tag{5.14}$$

Wird keine Verarbeitung im Teilbandbereich durchgeführt, gilt: $\hat{S}_\mu(k') = Y_\mu(k')$. Gleichung 5.13 kann dann in Gleichung 5.14 eingesetzt werden und man erhält für den Ausgang der Synthesefilterbank nach einigen Umformungen:

$$\begin{aligned}\hat{S}\left(e^{j\Omega}\right) = &\sum_{\mu=0}^{M-1} \frac{1}{r} Y\left(e^{j\Omega}\right) H\left(e^{j\left(\Omega-\frac{2\pi}{M}\mu\right)}\right) G\left(e^{j\left(\Omega-\frac{2\pi}{M}\mu\right)}\right)\cdots \\ &+\sum_{\mu=0}^{M-1} \frac{1}{r} \sum_{\xi=1}^{r-1} Y\left(e^{j\left(\Omega-\frac{2\pi}{r}\xi\right)}\right) H\left(e^{j\left(\frac{\Omega}{r}-\frac{2\pi}{r}\xi-\frac{2\pi}{M}\mu\right)}\right) G\left(e^{j\left(\Omega-\frac{2\pi}{M}\mu\right)}\right). \end{aligned} \tag{5.15}$$

In dieser Darstellung ist die mathematische Beschreibung des Spektrums des Syntheseausgangs in einen linearen Nutzanteil (erster Term) und einen Aliasing-Anteil (Doppelsummenterm) aufgespalten. Um den Einfluss von letzterem auf den Filterbankausgang gering zu halten, müssen demnach die Flankensteilheit und Sperrdämpfung der Prototyptiefpassfilter maximiert und die Unterabtastrate r daran angepasst werden [39]. Zur Veranschaulichung der Aliasingeffekte ist in Abbildung 5.5 der Filterbankausgang bei Anregung mit einem Rechteckimpuls dargestellt. Der Einfluss des Aliasing spiegelt sich in den Überschwingern zu Beginn und Ende des Impulses sowie im Ripple in dem Abschnitt konstanter Amplitude innerhalb des Impulses wieder.

An die Filter h_k und g_k werden offensichtlich ähnliche Anforderungen gestellt. Beide sollen möglichst gut einen vorgegebenen Spektralanteil ausfiltern. Daher liegt es auf der Hand, für Analyse- und Synthesestufe den gleichen Prototyptiefpass zu verwenden. Setzt man $g_k = h_{L_{\mathrm{PLP}}-k}$ so kann der Filterbankausgang folgendermaßen abgeschätzt werden [44]:

$$\hat{S}\left(e^{j\Omega}\right) \approx Y\left(e^{j\Omega}\right) \sum_{\mu=0}^{M-1} \frac{1}{r} \left|H\left(e^{j\left(\Omega-\frac{2\pi}{M}\mu\right)}\right)\right|^2 e^{-jL_{\mathrm{PLP}}\left(\Omega-\frac{2\pi\mu}{M}\right)}, \tag{5.16}$$

woraus sich die Gesamtverzögerung der Filterbank (Analyse- und Synthesestufe) näherungsweise zu L_{PLP} ergibt.

5.1.5 Design des Prototyptiefpasses

Für das Design des im vorgestellten Verfahren verwendeten Prototyptiefpassfilters h_k wurde der Ansatz gemäß [55] gewählt, welcher hier aber nur kurz skizziert

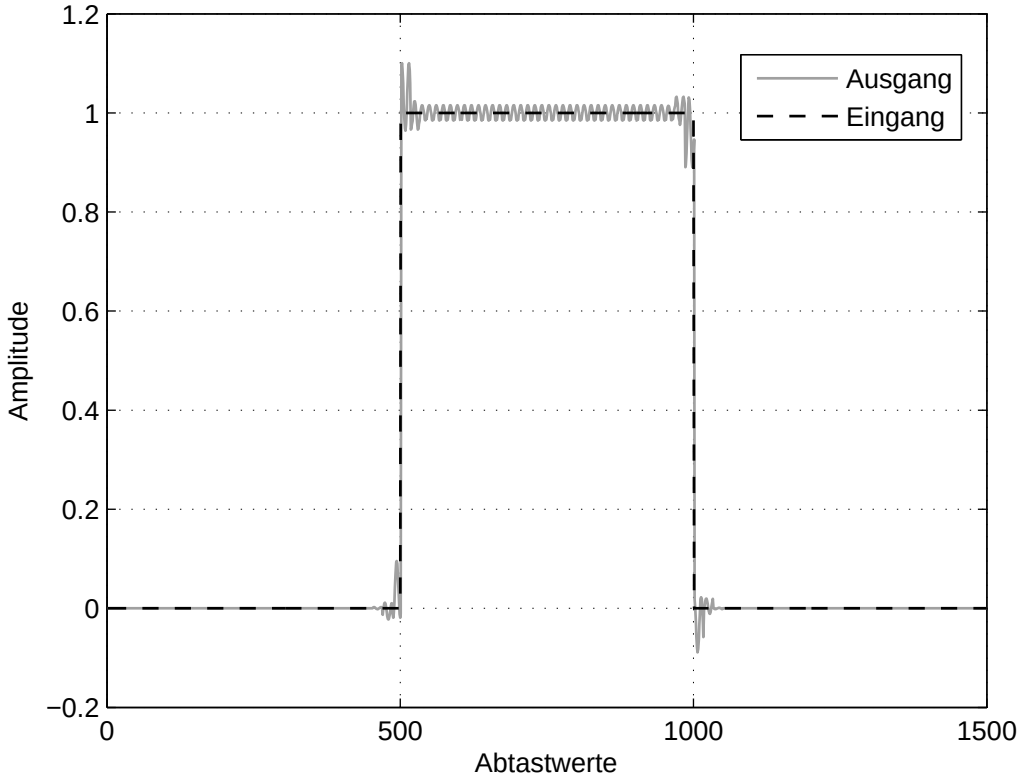

Abbildung 5.5: *Antwort einer Filterbank (Analyse und Synthese) bei Anregung mit einem Rechteckimpuls. Das Prototyptiefpassfilter hat die Länge* $L_{PLP} = 64$ *bei* $M = 16$ *Bändern und einer Unterabtastrate* $r = 12$.

wird. Das gesamte Entwurfsverfahren ist in [18, 44] beschrieben. Für $r = 1$ ergibt dieser Ansatz eine *perfekt-rekonstruierende* Filterbank. Darunter versteht man eine Filterbank, in deren Ausgang keinerlei Aliasingeffekte auftreten. Im vorliegenden Fall ist dies nur unter Weglassung der Unterabtastung (das heißt $r = 1$) gegeben, ansonsten treten mit steigendem r immer stärkere Aliasingartefakte auf. Daher muss bei dem Design eines Prototyptiefpasses die maximal benötigte Unterabtastung $r_{\max}$ berücksichtigt werden, da beispielsweise für große Werte (relativ zu M) von $r_{\max}$ andere Ansprüche an die Flankensteilheit resultieren als für kleine.

Zunächst wird aus Gleichung 5.16 eine Bedingung für den Frequenzgang $|H(e^{j\Omega})|^2$ von h_k abgeleitet:

$$\frac{1}{r}\left|H\left(e^{j\left(\Omega-\frac{2\pi}{M}\mu\right)}\right)\right|^2 \;=\; 1 \qquad \text{für } \mu = 0,\ldots,M-1, \tag{5.17}$$

welche in eine Bedingung für die Autokorrelationsfunktion des Prototyptiefpasses umgewandelt wird. Um diese zu erfüllen, wird als Autokorrelierte das Produkt zweier Funktionen angesetzt, für welche sich die Wahl der si-Funktion einerseits und der Tschebyscheff-Funktion andererseits bewährt hat. Mit dem Algorithmus nach Boite und Leich [6] wird schließlich die Impulsantwort des Prototyptiefpassfilters aus der Autokorrelierten synthetisiert.

Zur Kontrolle, ob der berechnete Prototyptiefpass die gestellten Anforderungen bezüglich Sperrdämpfung und Flankensteilheit erfüllt, kann die bandabhängige Aliasingkomponente $A_\mu(e^{j\Omega})$ aus Gleichung 5.15 bestimmt werden:

$$A_\mu(e^{j\Omega}) \;=\; \frac{1}{r}\sum_{\xi=1}^{r-1} Y\left(e^{j\left(\Omega-\frac{2\pi}{r}\xi\right)}\right) H\left(e^{j\left(\frac{\Omega}{r}-\frac{2\pi}{r}\xi-\frac{2\pi}{M}\mu\right)}\right) G\left(e^{j\left(\Omega-\frac{2\pi}{M}\mu\right)}\right). \quad (5.18)$$

Als Maß für die Güte kann daraus die in [55] vorgeschlagene Antwort des Filterbanksystems auf den Einheitsimpuls $y(k)=\gamma\delta_{\mathrm{K}}(k)$ verwendet werden. Dabei bezeichnet γ wiederum (siehe Abschnitt 3.4.2) einen Normierungsfaktor mit Betrag Eins und gleicher Dimension wie $y(k)$. Für dieses Testsignal verschwindet der vom Eingangssignal abhängige Teil aus Gleichung 5.18 und man erhält:

$$A_\mu(e^{j\Omega}) \;=\; \frac{1}{r}\sum_{\xi=1}^{r-1} \gamma H\left(e^{j\left(\frac{\Omega}{r}-\frac{2\pi}{r}\xi-\frac{2\pi}{M}\mu\right)}\right) G\left(e^{j\left(\Omega-\frac{2\pi}{M}\mu\right)}\right). \quad (5.19)$$

Im vorgeschlagenen Verfahren wurde eine Filterbankstruktur mit $M=16$ Teilbändern und einem Prototyptiefpassfilter der Länge $L_{\mathrm{PLP}}=64$ verwendet. Letzteres ist für eine maximale Unterabtastung von $r_{\max}=12$ ausgelegt und in Abbildung 5.6 dargestellt.

Die Wahl der Unterabtastrate r hat zudem Einfluss auf die effektive Bandbreite B_{FB} der einzelnen Teilbänder. Für diese gilt bei idealer Filterung:

$$B_{\mathrm{FB}} \;=\; \frac{f_s}{r}. \quad (5.20)$$

In Tabelle 5.1 sind die effektiven Bandbreiten der einzelnen Teilbänder der in dieser Arbeit verwendeten Filterbank für verschiedene Unterabtastraten r aufgelistet. Dabei ist zu beachten, dass für die reellwertigen Frequenzbänder, das sind das erste ($\mu=0$) und das neunte Band ($\mu=M/2$), aufgrund deren Symmetrie nur die halbe Bandbreite angegeben wird. Im Folgenden werden die für die Implementierung relevanten ersten $M/2+1$ Teilbänder der Einfachheit halber mit erstes, zweites, usw. Band bezeichnet, um Bezeichnungen wie *nulltes Teilband* zu vermeiden.

In der Anwendung, also bei nichtidealer Filterung, verbreitern sich die einzelnen Teilbänder noch weiter, da unter- bzw. oberhalb der jeweiligen Grenzfrequenzen die Sperrdämpfung aufgrund der endlichen Filterlänge L_{PLP} nicht sofort hinreichend große Werte annimmt, wodurch das zuvor beschriebene Aliasing entsteht.

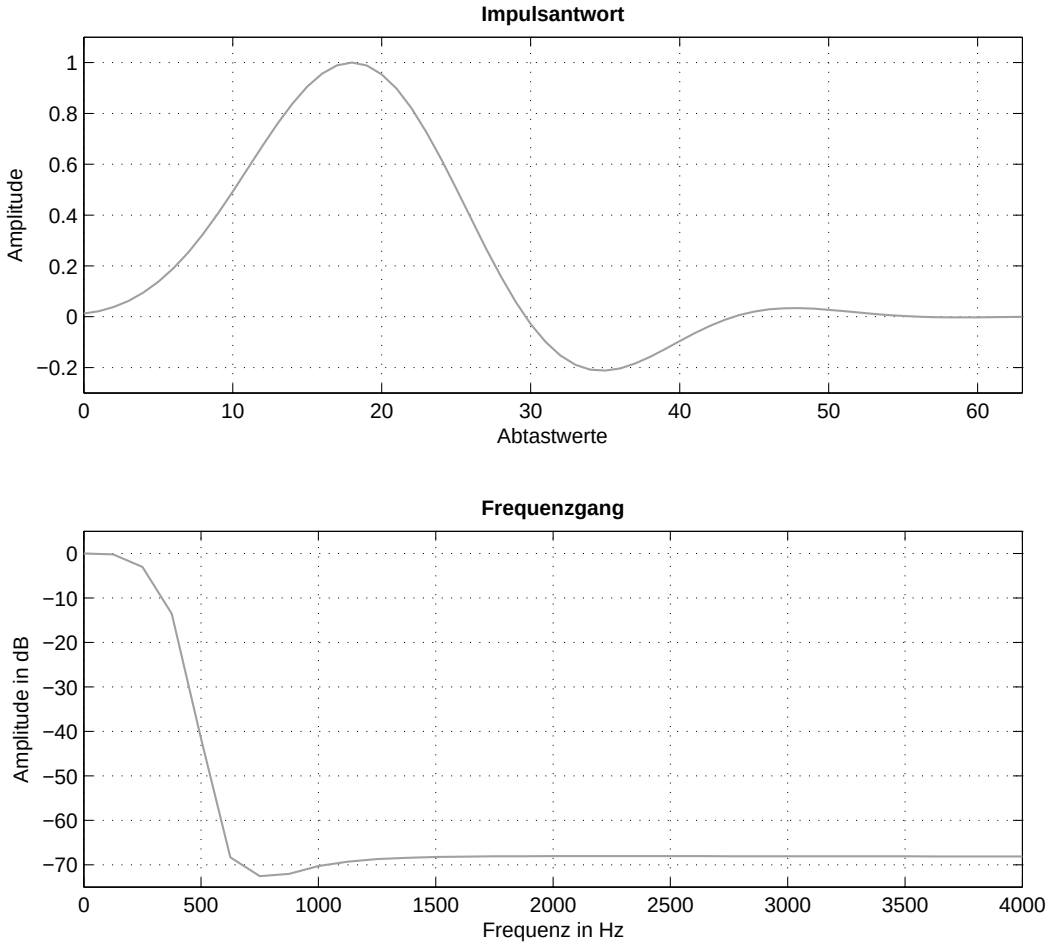

Abbildung 5.6: *Impulsantwort und Frequenzgang des verwendeten Prototyptiefpassfilters.*

5.2 Parametrierung der zeitinvarianten Größen

Im folgenden Abschnitt wird die Wahl der noch nicht festgelegten, über der Zeit konstanten Parameter diskutiert. Dabei wird insbesondere auf die Wahl der Ordnung für das Sprachmodell eingegangen. Da jetzt explizit Teilbandsignale betrachtet werden, wird als Zeitindex k' verwendet (siehe Abschnitt 5.1.3), um anzuzeigen, dass unterabgetastete Signale vorliegen.

Tabelle 5.1: *Untere und obere Grenzfrequenz in Hertz der jeweiligen Teilbänder bei $M = 16$, $f_s = 8$ kHz und reellen Eingangssignalen für unterschiedliche r.*

Teilband	**r = 16**	**r = 12**	**r = 10**
1	0-250	0-333	0-400
2	250-750	167-833	100-900
3	750-1250	667-1333	600-1400
4	1250-1750	1167-1833	1100-1900
5	1750-2250	1667-2333	1600-2400
6	2250-2750	2167-2833	2100-2900
7	2750-3250	2667-3333	2600-3400
8	3250-3750	3167-3833	3100-3900
9	3750-4000	3667-4000	3600-4000

5.2.1 Wahl der Ordnungen

Durch die Implementierung des in Kapitel 3 vorgestellten Kalman-Filter-Algorithmus' in einer Filterbankstruktur (siehe Abschnitte 5.1.2 und 5.1.3) besteht die prinzipielle Möglichkeit, die Ordnung p des Sprachmodells sowie die Ordnung q der Geräuschmodelle für jedes Teilband verschieden zu wählen. Das konnte bis jetzt vernachlässigt werden, da stets nur ein Frequenzband betrachtet wurde. Genaugenommen stellen die Ordnungen des AR-Modells aber zeitinvariante Funktionen des Bandindex' μ dar, also p_μ und q_μ.

Zudem muss berücksichtigt werden, dass der numerische Aufwand des Kalman-Filters kubisch sowohl mit der Ordnung des Sprach- als auch des Geräuschmodells wächst (siehe Abschnitt 3.5).

Ordnung des Sprachmodells

Die Wahl der für jedes Teilband verschiedenen Sprachmodellordnungen wurde in [39] über den mittleren Gehalt an Nutzsignalenergie der einzelnen Bänder begründet. Die zugrunde liegende Idee ist dabei, in Bändern mit viel Nutzsignalenergie relativ hohe Ordnungen für das AR-Modell zu wählen, da hier gut an das Modell angepasst werden kann und man dort an einer genauen Modellierung besonders interessiert ist (wegen des guten SNRs). Dagegen reichen für Teilbänder mit niedrigem SNR wenige Koeffizienten aus, um das Spektrum ausreichend zu charakterisieren. Eine solche Energieverteilung über die Teilbänder eines aufgenommenen Signals ist in Abbildung 5.7 dargestellt.

Im untersten Band ist nur wenig Nutzsignal enthalten. Dies war zu erwarten,

da im Bereich unterhalb der Sprachgrundfrequenz (siehe Abschnitt 2.3) keine Signalenergie vorhanden ist und dieser Bereich einen signifikanten Teil des Bandes einnimmt. Für die übrigen Teilbänder ergibt sich der bereits in Kapitel 2 beschriebene Verlauf mit fallender Signalenergie bei größer werdendem μ. Anhand dieser Energieverteilung wird bestimmt, welches Band mit vielen und welches mit wenigen AR-Koeffizienten parametrisiert wird.

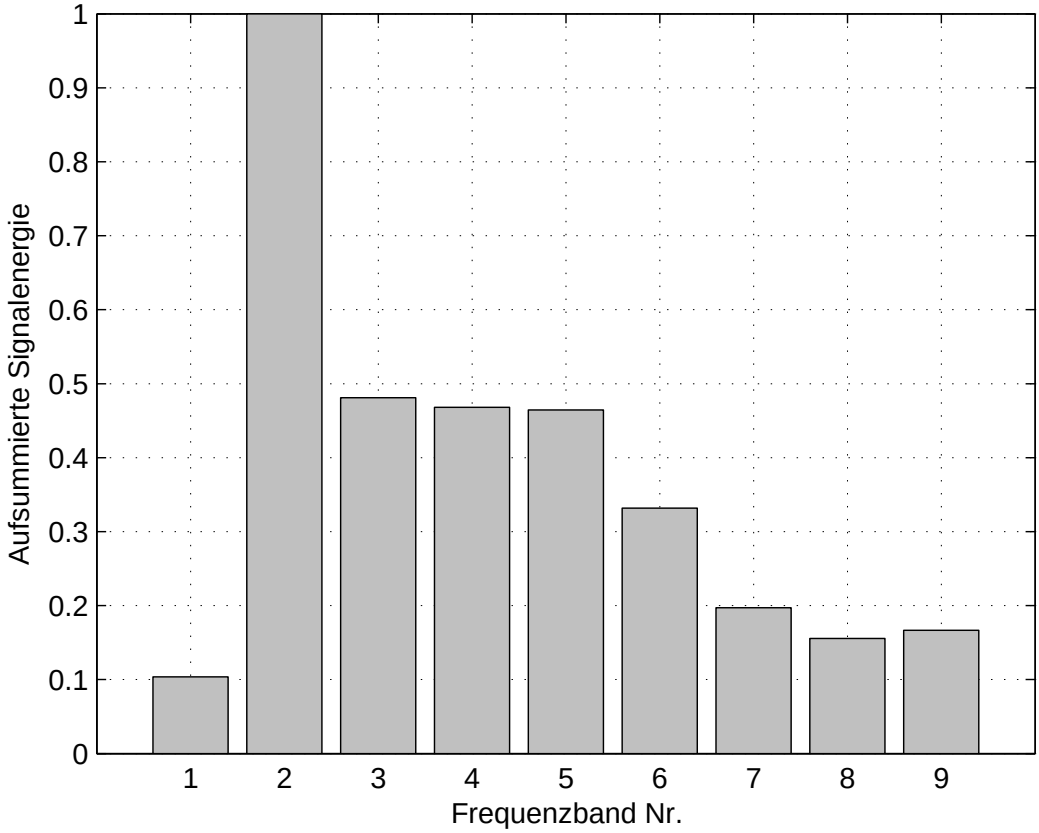

Abbildung 5.7: *Aufsummierte und normierte Signalenergie gemittelt über verschiedene Sprachsignale für jedes Teilband.*

Diese relative Zuordnung ist alleine noch nicht aussagekräftig genug, da daraus nicht hervorgeht, wie viele Koeffizienten pro Teilband genau Verwendung finden sollen. Dazu kann man zunächst eine theoretische Abschätzung durchführen: Ein tiefe Männerstimme hat eine Sprachgrundfrequenz von etwa 100 Hz (siehe Abschnitt 2.1.3). Unter der Annahme einer ideal harmonischen Struktur und $f_s = 8000$ Hz ergeben sich zusätzlich 39 Oberwellen, was ca. vier Oberwellen pro Teilband entspricht (im ersten und $M/2+1$-ten entsprechend weniger).

Diese Abschätzung setzt allerdings voraus, dass die Kanäle ideal getrennt werden,

was bei dem vorliegenden Prototyptiefpassfilter nicht gegeben ist. In Abbildung 5.8 wurde das Sprachsignal eines kurzen stimmhaften Abschnitts zehnmal mit der Analysestufe der in dieser Arbeit verwendeten Filterbank verarbeitet. Beim ersten Mal wurden anschließend alle Teilbandsignale der Synthesestufe zugeführt, beim zweiten Mal nur das des ersten Bands, beim dritten Mal nur das des zweiten Band, etc. und die resultierenden Zeitsignale als Spektrogramm dargestellt.

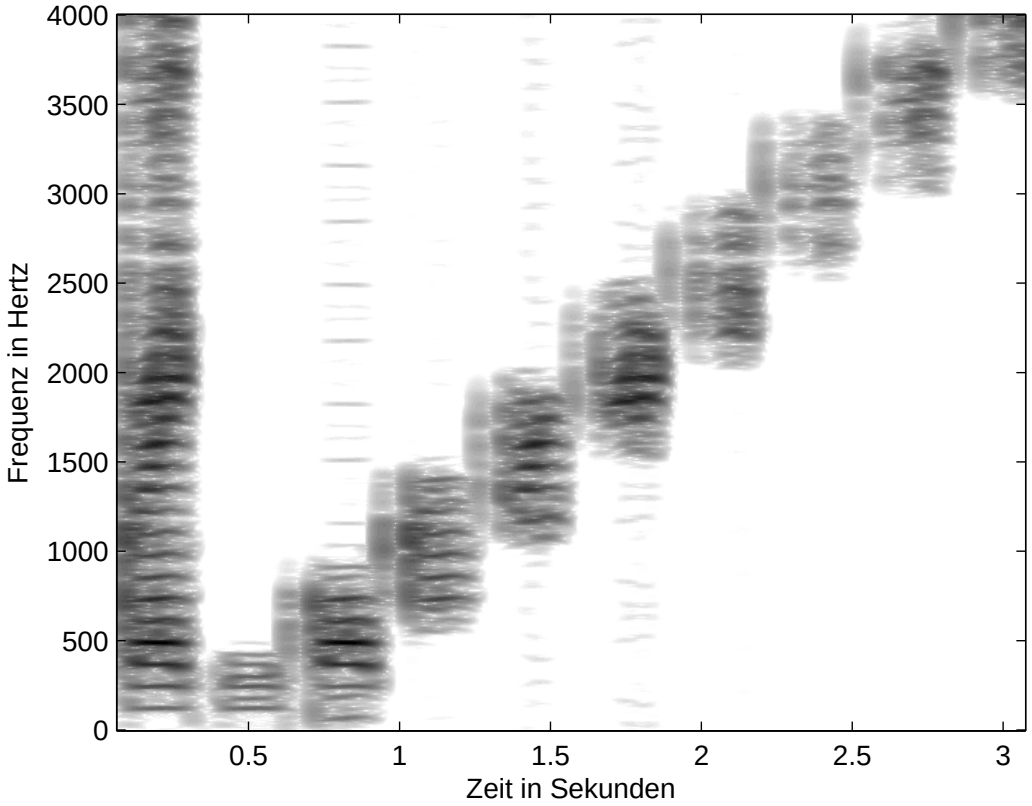

Abbildung 5.8: *Darstellung des Vollbandsignals und der einzelnen Teilbandsignale bei gleichem Eingang.*

Zwei Dinge sind eindeutig zu erkennen. Zum einen überlappen sich die einzelnen Teilbänder aufgrund der nicht idealen Prototyptiefpassfilter und der Unterabtastrate von $r = 10$. Zum anderen tritt Aliasing in den Abschnitten außerhalb des jeweiligen Frequenzbands auf. Durch diese Verbreiterung der Teilbänder fallen bis zu sechs Pitchkomponenten in ein Band.

Auf Grundlage beider Überlegungen werden in dieser Arbeit die in Tabelle 5.2 aufgeführten Sprachmodellordnungen verwendet.

Für das erste und das $M/2+1$-te Teilband sollten aufgrund der vorliegenden

Symmetrie stets gerade Ordnungen gewählt werden. Da diese Bänder nur mit ihrer einseitigen Bandbreite dargestellt sind, muss darüber hinaus beachtet werden, dass bei beispielsweise zwei vorhandenen Pitchkomponenten die doppelte Ordnung, also $p=4$, verwendet werden muss.

Tabelle 5.2: *Verwendete Ordnung p für das AR-Sprachmodell in Abhängigkeit des Teilbandindex μ.*

Teilband μ	Frequenzbereich	Sprachmodellordnung p
1	0-250	4
2	250-750	6
3	750-1250	6
4	1250-1750	6
5	1750-2250	6
6	2250-2750	5
7	2750-3250	4
8	3250-3750	4
9	3750-4000	4

Ordnung der Geräuschmodelle

Da das Geräusch keine ausgeprägte Feinstruktur aufweist, kann es generell mit weniger AR-Koeffizienten als das Sprachmodell parametrisiert werden. Von diesem Standpunkt aus genügen zwei Koeffizienten pro Teilband.

Da q gleichzeitig die Länge des Geräuschvektors $\mathbf{n}_i(k')$ und damit auch die Länge der adaptiven Filter $g_{i_1 i_2}(k')$ darstellt, erscheint eine Vergrößerung auf $q > 2$ zunächst sinnvoll, um die Wirksamkeit dieser Filter mit wachsender Filterlänge zu verbessern. Vergleichstests haben allerdings gezeigt, dass der erzielbare Gewinn für das Gesamtsystem an dieser Stelle nur sehr gering ist, so dass die Erhöhung der Geräuschmodellordnung auf $q>2$ verworfen wurde.

5.2.2 Wahl der übrigen Konstanten

Unterabtastrate und Blocklängen im Teilband

In dem einkanaligen Verfahren aus [39] wurde vorgeschlagen, bei einer Unterabtastrate von $r = 12$ und gleichem Prototyptiefpassfilter eine Teilbandblocklänge von $L_{\mathrm{AR}} = 32$ zu wählen, welche bei der Berechnung des Periodogramms in der DAKF durch Zero-Padding auf die Länge $L'_{\mathrm{AR}} = 64$ verlängert wird.

Wie in Abschnitt 4.3.1 aufgeführt, ist es ausreichend, bei einer maximalen Ordnung von $p_{\max}=6$ nur sieben Nullen anzuhängen. Um die für die FFT günstige Struktur aufrecht zu erhalten, müsste die Blocklänge in diesem Fall auf $L_{\mathrm{AR}}=57$ erweitert werden. Dies entspräche bei $r=12$ einem Zeitfenster von 86 ms, was deutlich über dem für die Stationaritätsannahme aus Abschnitt 2.1.1 gültigen Zeitinvervall liegt ($L_{\mathrm{AR}}=32$ entspricht bereits 48 ms).

Eine möglich Abhilfe besteht darin, die Blocklänge auf $L_{\mathrm{AR}}=25$ zu verkürzen, so dass nach Anfügen der minimal notwendigen Nullen wieder eine Zweierpotenz vorliegt ($L'_{\mathrm{AR}}=32$). Diese Verkürzung bedeutet allerdings den Verzicht auf 22% des Datenumfangs des ursprünglichen Blocks, wodurch sich das korrespondierende Zeitfenster auf 38 ms verkürzt, was hinsichtlich der Stationaritätsannahme noch akzeptabel ist, aber aufgrund des geringeren Datenumfangs eine etwas schlechtere Parameterschätzung bewirkt.

Eine weitere Möglichkeit besteht darin, den Block der Länge $L'_{\mathrm{AR}}=64$ mit mehr Daten zu füllen, anstatt ihn auf $L'_{\mathrm{AR}}=32$ zu verkürzen. Dazu wird die Unterabtastrate auf $r=10$ reduziert und die ursprüngliche Blocklänge auf $L_{\mathrm{AR}}=38$ verlängert. Das korrespondierende Zeitfenster bleibt durch diese Maßnahme unverändert 48 ms lang, gleichzeitig wächst der Datenumfang jedes Blocks um 19%. Dies wird erkauft durch Mehraufwand bei der Teilbandverarbeitung, da die Unterabtastrate verringert wurde. Diese Reduktion bringt aber auch an dieser Stelle einen Vorteil. Verarbeitet man ein ungestörtes Sprachsignal mittels der hier verwendeten Analyse- und Synthesefilterbank bei $r=12$, so können im Sprachsignal des Ausgangs bereits durch Aliasing verursachte Sprachverzerrungen gehört werden, welche bei $r=10$ noch nicht hörbar sind. Bei Verwendung innerhalb einer Geräuschreduktionsstruktur werden diese durch das Restrauschen größtenteils maskiert, weshalb dort $r=12$ benutzt werden kann. Eine Reduktion der Unterabtastrate auf $r=10$ verhindert somit durch die Filterbank verursachte, hörbare Verzerrungen, weshalb dieses Verarbeitungsschema in der vorliegenden Arbeit verwendet wird. Der entstehende numerische Mehraufwand ist bei den heute verfügbaren Mikroprozessorleistungen vertretbar.

Blockversatz im Teilband

Am Ausgang des Kalman-Filters wird aus dem Schätzwert des Zustandsvektors $\hat{\mathbf{x}}(k'|k')$ pro Zyklus nur ein Wert extrahiert – üblicherweise $\hat{s}(k')$. Die Berechnung wird somit Abtastwert für Abtastwert durchgeführt. Dem Gegenüber steht die Schätzung der Parameter, die auf der Verarbeitung von Blöcken der Länge L_{AR} beruht. Diese müssen so gelegt werden, dass der interessierende Zeitpunkt in der Blockmitte liegt. Dafür muss eine künstliche Verzögerung um $L_{\mathrm{AR}}/2-1$ in den Signalpfad hinzugefügt werden, um diese Nicht-Kausalität aufzulösen.

Darüber hinaus ist es nicht notwendig, für jeden Zeitschritt den Block um Eins weiterzuschieben. Vielmehr kann ein Block für mehrere Verarbeitungszyklen benutzt werden, da Kurzzeitstationarität gegeben ist. Die zu verarbeitenden Abtastwerte werden in diesem Fall um die Mitte des Blocks verteilt. Tests haben ergeben, dass bei gegebener Abtastrate von $r = 10$ ein Block für bis zu sechs aufeinander folgende Zyklen verwendet werden kann, bevor er um wiederum $Q = 6$ Abtastwerte weitergeschoben wird. Dieses Ergebnis entspricht von der zeitlichen Verzögerung dem in [39] bei einer Unterabtastrate von $r = 12$ gefundenen Wert $Q = 5$.

5.2.3 Betrachtung der Gesamtverzögerung

Abschließend soll die durch das Verfahren verursachte Verzögerung betrachtet werden. Für die Anwendung im Mobilfunkstandard GSM wurde durch das *European Telecommunications Standard Institute (ETSI)* eine maximale Verzögerung von 39 ms definiert [13], die auch hier als Obergrenze dienen soll.

Eine Verzögerung tritt an zwei Stellen auf:

- Filterbank,
- AR-Modell Schätzung.

Durch die Analyse- und Synthesestufe der Filterbank entsteht eine Verzögerung von $L_{\mathrm{PLP}} = 64$ Abtastwerten bei einer Abtastrate von $f_s = 8000$ Hz, was einem zeitlichen Versatz von 8 ms entspricht. Der Ausgleich der im vorangegangenen Abschnitt diskutierten Nicht-Kausalität verursacht darüber hinaus eine Verzögerung um weitere $L_{\mathrm{AR}}/2 = 19$ Abtastwerte bei einer Unterabtastrate von $r = 10$. Dies entspricht aufgerundet 24 ms. Die Gesamtverzögerung bleibt mit 32 ms damit unterhalb den von dem ETSI geforderten 39 ms.

Die Differenz von 7 ms kann darüber hinaus genutzt werden, um die Schätzung der Sprachkomponente durch Nachglättung (siehe Abschnitt 3.4.2) zu verbessern. Bei einer Unterabtastrate von $r = 10$ entspricht ein zusätzlicher Zeitschritt einer Verzögerung von $1,25$ ms. Bei den zur Verfügung stehenden 7 ms bedeutet dies, dass maximal um $5 \cdot 1,25 = 6,25$ ms zusätzlich verzögert werden kann. Mit $p_{\max} = 6$ entspricht das der maximal möglichen Nachglättung. Im Sinne einer möglichst kurzen Signalverzögerung und um die Handhabung der Teilbänder mit $p < p_{\max}$ zu vereinfachen, wird in dieser Arbeit von $p_{\min} = 4$ ausgegangen und eine zusätzliche Verzögerung von drei Abtastzyklen verwendet. Dadurch erhöht sich die Gesamtverzögerung um $3,75$ ms auf aufgerundet 36 ms.

Kapitel 6

Bewertung der Simulationsergebnisse

In diesem Kapitel werden Simulationsergebnisse vorgestellt und diskutiert. Im ersten Teil wird dazu zunächst die Leistungsfähigkeit der mehrkanaligen Methoden zur Schätzung der AR-Parameter aus Abschnitt 4.3.2 verglichen und bewertet. Zusätzlich wird der Einfluss verschiedener Faktoren wie zum Beispiel der Kanalanzahl und des SNRs aufgezeigt.

Der zweite Teil beschäftigt sich mit der Leistungsfähigkeit des Gesamtsystems, also der Güte der Geräuschreduktion, wiederum für verschiedene Konfigurationen und im Vergleich zu anderen Strukturen, beispielsweise der Kombination Beamformer und einkanaligem Kalman-Filter. Zum Schluss wird im dritten Teil ein Fazit gezogen.

6.1 Leistungsfähigkeit der mehrkanaligen Schätzerverfahren

In diesem Teil des Kapitels werden die Simulationsergebnisse der mehrkanaligen AR-Parameter-Schätzmethoden vorgestellt. Alle gezeigten Diagramme zeigen für die Situation typische Verläufe und stellen keine Spezialfälle dar. Dargestellt sind jeweils diejenigen Signalspektren, die sich aus den geschätzten AR-Parametern ergeben. Dabei ist die Schätzung der *Methode 1* (Kanalmittelung über Reflexionskoeffizienten) stets schwarz, die der *Methode 2* (SNR gewichtete Kanalmittelung über Reflexionskoeffizienten) gestrichelt schwarz, die der *Methode 3* (Kanalmittelung über Periodogramme) gepunktet schwarz und das Referenzsignal, das ist die Schätzung bei unverrauschtem Eingang, grau abgebildet.

6.1.1 Vergleich der Verfahren untereinander

In den Abbildungen 6.1 und 6.2 sind die geschätzten Spektren pro Teilband eines stimmhaften Lautes für $N = 4$ Mikrofone und 15 dB SNR dargestellt.

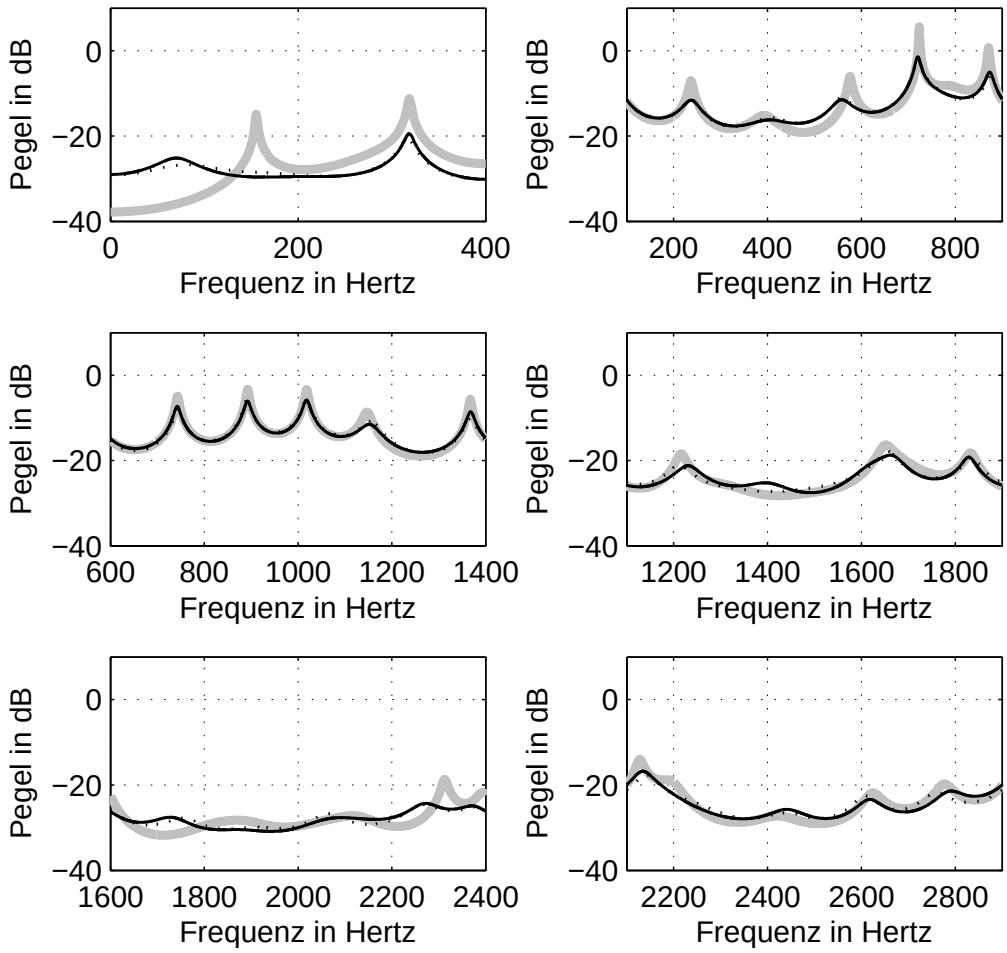

Abbildung 6.1: *Von links oben nach rechts unten: Spektralschätzung eines stimmhaften Lautes in den Teilbändern 1 bis 6. Fortsetzung in Abbildung 6.2.*

Anhand dieser sollen nun die in 4.3.2 vorgestellten Methoden verglichen werden.

Zuerst fällt auf, dass der Verlauf der geschätzten Spektren von *Methode 1* und *Methode 2* nahezu identisch ist. Dies bedeutet, dass die Signale am Ausgang der Mikrofone hinsichtlich ihrer Verwendbarkeit in der AR-Parameterschätzung alle

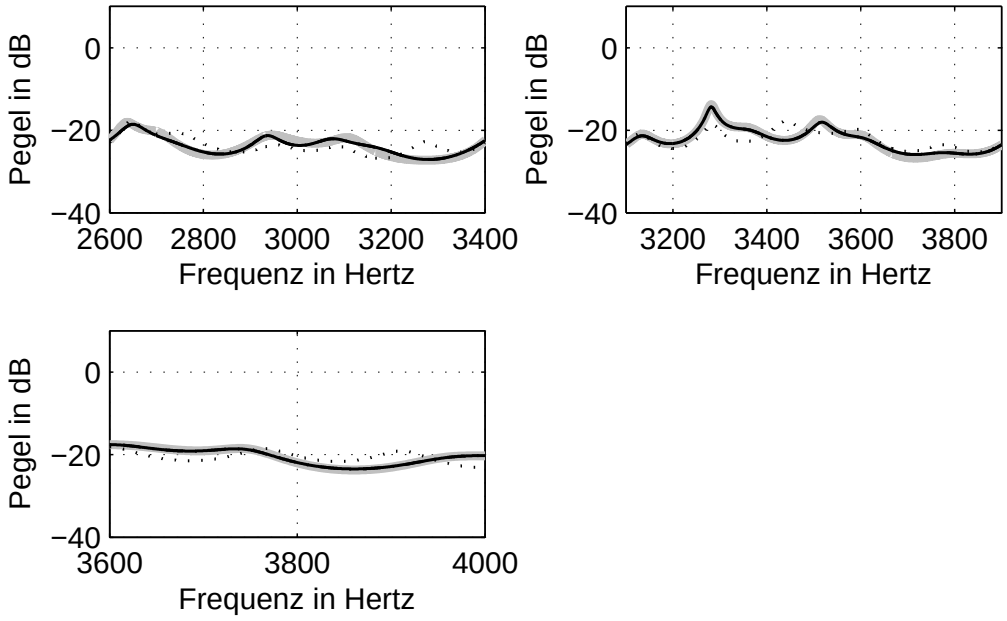

Abbildung 6.2: *Von links oben nach rechts unten: Spektralschätzung eines stimmhaften Lautes in den Teilbändern 7 bis 9. Fortsetzung von Abbildung 6.1.*

etwa gleich gut geeignet sind. Eine Gewichtung mit dem geschätzten SNR des jeweiligen Kanals bringt in diesem Fall keine weitere Verbesserung. Für die hier abgebildeten Spektren wurden vier am Rückspiegel montierte Mikrofone (siehe Abbildung 2.10) verwendet. Selbst bei Wahl der Mikrofone mit dem größten Laufzeitunterschied bleiben die Unterschiede kaum sichtbar, so dass der zwar nur geringe Mehraufwand von *Methode 2* gegenüber *Methode 1* nicht gerechtfertigt werden kann und *Methode 2* im Folgenden nicht mehr von *Methode 1* unterschieden wird. Zu einem vergleichbaren Ergebnis kommt man, wenn *Methode 3* mit der SNR-Gewichtung der Kanäle kombiniert wird. Auch hier sind die resultierenden Graphen nahezu identisch, so dass der Mehraufwand nicht lohnt.

Für dieses überraschende Ergebnis gibt es zwei Gründe. Zum einen sind die sinnvoll möglichen Mikrofonanordnungen im Kraftfahrzeug beschränkt, da sich die Mikrofone stets im Raumsegment vor dem Fahrer befinden und üblicherweise in einem linearen Array angeordnet sind. Dadurch ergeben sich keine allzu großen Laufzeit- und somit Pegelunterschiede. Zum anderen wurden für die Aufnahmen der Audiodaten sehr hochwertige Mikrofone verwendet, die eine sehr gute Übertragungscharakteristik sowie eine sehr geringe Streuung untereinander aufweisen. Dies ist für Serienfahrzeuge nicht mehr gegeben, so dass selbst dicht benachbarte Mikrofone durchaus signifikante Unterschiede im SNR aufweisen können. Daher

ist davon auszugehen, dass unter weniger idealen Bedingungen als sie für die verwendeten Audiodaten vorlagen, *Methode 2* durchaus bessere Ergebnisse liefern kann als *Methode 1*, weshalb auf deren Darstellung im Rahmen dieser Arbeit nicht verzichtet wurde.

Im ersten Teilband ist gut zu erkennen, dass die auf der DAKF basierten Algorithmen dazu tendieren, die Frequenz der untersten Pitchkomponente – also der Sprachgrundfrequenz – zu unterschätzen. Dieses Problem kann mittels dem in [39] beschriebenen Verfahren, welches speziell für das unterste Teilband vorgeschlagen wurde und dort die Lage der Pole korrigiert, hier aber nicht benutzt wird, behoben werden.

In den übrigen Teilbändern sind die Verläufe der einzelnen Methoden ähnlich, mit leichten Vorteilen für *Methode 1* gegenüber *Methode 3* besonders in den oberen Teilbändern, wo stellenweise dem wahren Verlauf überhaupt nicht mehr gefolgt wird. Man kann daher sagen, dass unter den hier gegebenen typischen Voraussetzungen, das ist insbesondere ein mittleres SNR, *Methode 1* die beste Wahl darstellt (für sehr schlechtes SNR siehe Abschnitt 6.1.2).

6.1.2 Einfluss des SNRs

In Abbildung 6.3 sind die geschätzten Sprachsignalspektren des selben stimmhaften Sprachabschnitts wie in Abschnitt 6.1.1 für verschiedene Signal-zu-Geräuschverhältnisse (SNR) bei konstanter Kanalanzahl $N = 4$ dargestellt. Dabei entsprechen die Diagramme für 10 und 20 dB SNR den im Kraftfahrzeug typischerweise anzutreffenden Werten, während 0 und 30 dB SNR bereits als Extremwerte bezeichnet werden können.

Zunächst kann der zu erwartende Verlauf beobachtet werden. Die Schätzungen verbessern sich mit steigendem SNR. Können bei 0 dB die beiden rechten Pitchkomponenten noch nicht getrennt werden, ist dies bei 10 dB bereits möglich. Allerdings weist die geschätzte Frequenz der schwächsten Pitchkomponente noch einen Fehler auf. Dieser verkleinert sich mit weiter steigendem SNR, verschwindet jedoch selbst bei 30 dB SNR nicht vollständig.

Besonders zu beachten ist, dass *Methode 3* bei sehr schlechtem SNR (0 dB) im Gegensatz zu *Methode 1* noch in der Lage ist, die zweite abgebildete Pitchkomponente zu detektieren. Obwohl wie in Abschnitt 6.1.1 festgestellt, *Methode 1* bei typischem SNR prinzipiell immer bessere Resultate liefert als *Methode 3*, kehrt sich dieser Zusammenhang für sehr niedriges SNR um. In diesem Fall wirkt sich die Varianzverbesserung bei der Periodogrammmittelung (*Methode 3*) stärker aus als die Mittelung der Reflexionskoeffizientensätze (*Methode 1*), da letztere aus stark verrauschten Eingangsdaten berechnet wurden.

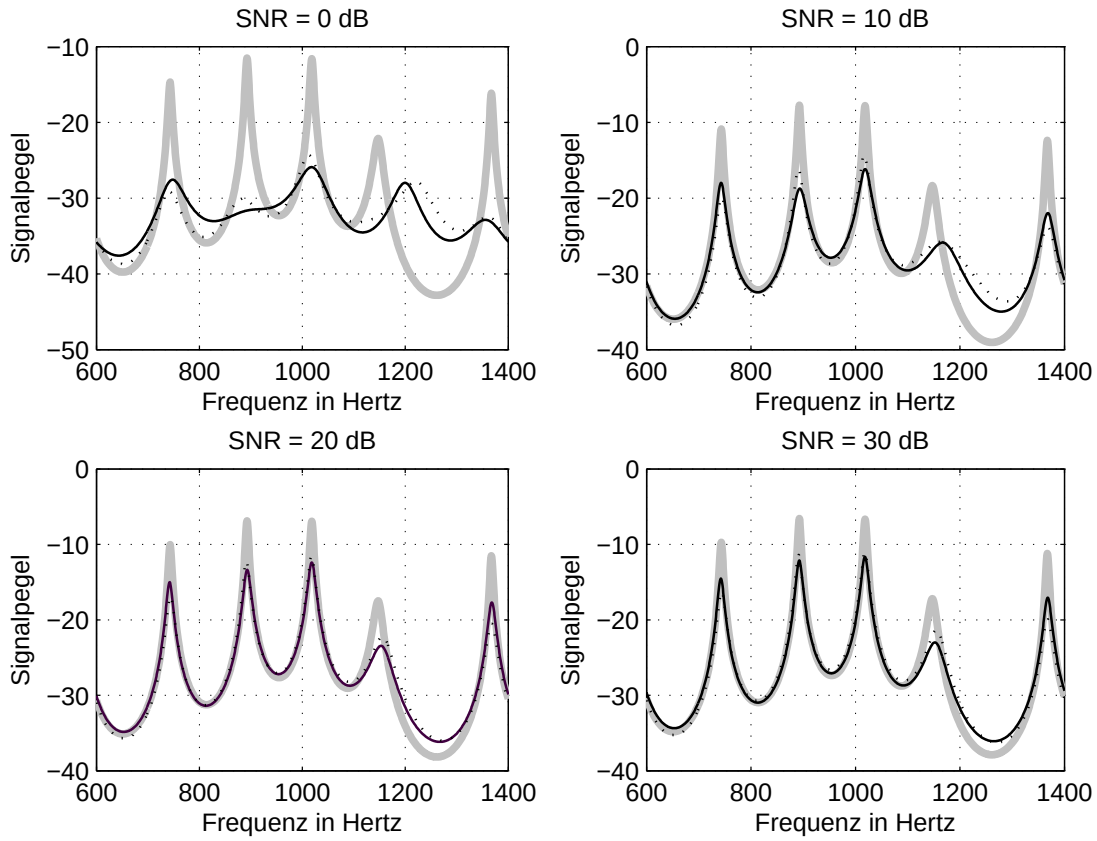

Abbildung 6.3: *Einfluss des SNRs auf die Güte der AR-Koeffizienten-Schätzung im dritten Teilband bei $N = 4$ Mikrofonen. Beginnend mit 0 dB oben links ansteigend bis 30 dB unten rechts.*

6.1.3 Einfluss der Kanalanzahl

In den Abbildungen 6.4 und 6.5 ist der bereits mehrfach dargestellte stimmhafte Sprachabschnitt für verschiedene Kanalanzahlen N sowie ein SNR von 15 dB bzw. 5 dB dargestellt.

Für den einkanaligen Fall liefern *Methode 1* und *Methode 3* die gleichen Ergebnisse, da hier keine Mittelung, die bei beiden Methoden unterschiedlich durchgeführt wird, notwendig ist.

Mit steigender Kanalanzahl N steigt die Schätzgüte bei einem SNR von 15 dB nur geringfügig an. Lediglich die Frequenzschätzung der schwächsten Pitchkomponente verbessert sich merklich gegenüber dem Fall mit wenigen Kanälen. Eine hohe Kanal- bzw. Mikrofonanzahl erhöht hier hauptsächlich die Robustheit der Schätzung.

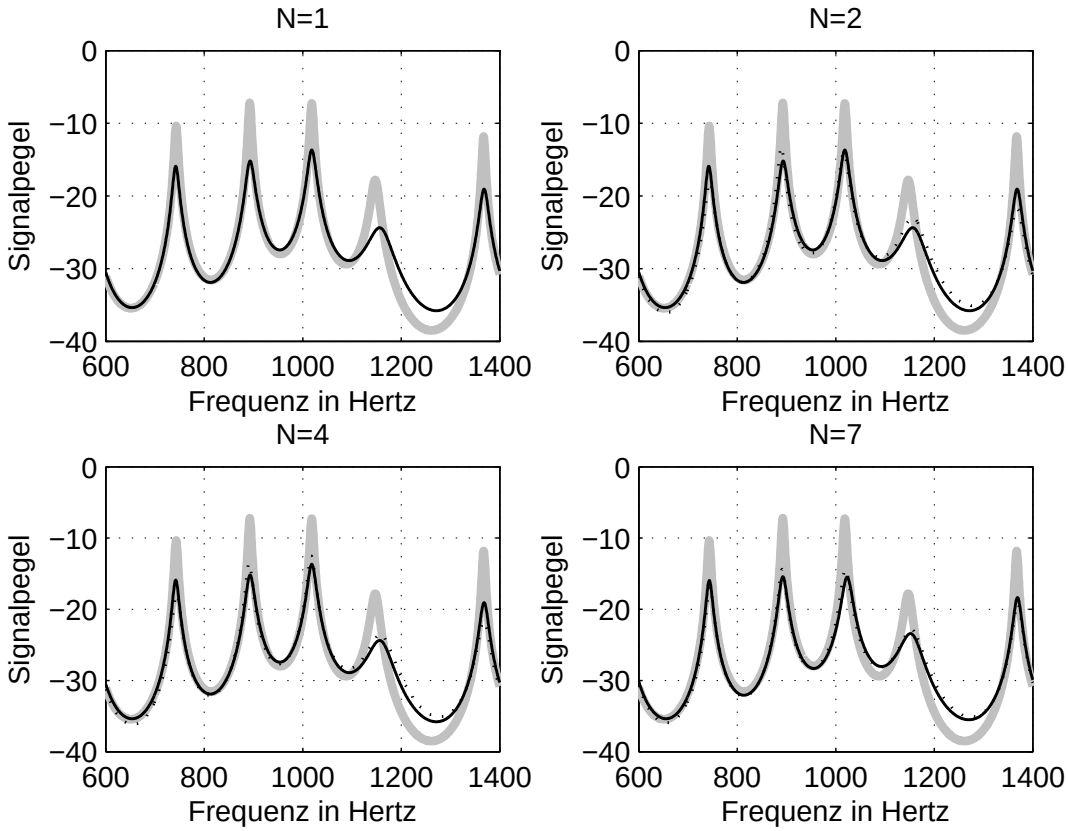

Abbildung 6.4: *Einfluss der Kanalanzahl auf die Güte der AR-Koeffizienten-Schätzung im dritten Teilband bei einem SNR von 15 dB.*

Liegt ein relativ niedriges SNR vor, so bewirkt eine Vergrößerung der Kanalanzahl eine merkliche Verbesserung der gesamten Schätzung. Wiederum kann beobachtet werden, dass *Methode 3* in diesem Fall – also bei schlechtem SNR – bessere Ergebnisse liefert als *Methode 1* (siehe Abbildung 6.5).

6.1.4 Einfluss der Sprachmodellordnung

In Abbildung 6.6 ist das bereits bekannte stimmhafte Testsignal für verschiedene Ordnungen des Sprachmodells dargestellt. Anhand des bei ungestörtem Sprachsignal berechneten Referenzspektrums kann abgelesen werden, dass die notwendige Ordnung in diesem Beispiel $p_{\mathrm{ref}} = 5$ beträgt. Dargestellt sind sowohl kleine (Differenz ± 1) als auch große (Differenz ± 3) Unter- und Überschätzungen der Sprachmodellordnung.

Generell reagieren die hier vorgestellten Algorithmen auf zu große Ordnungen

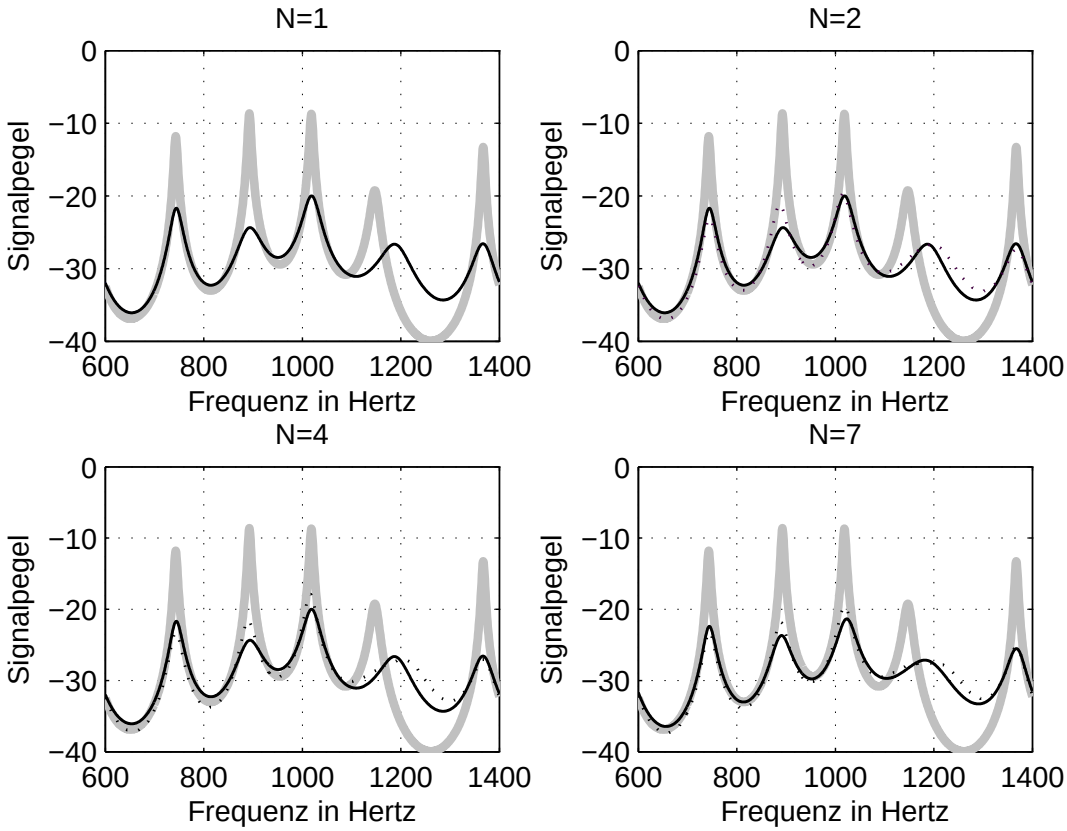

Abbildung 6.5: *Einfluss der Kanalanzahl auf die Güte der AR-Koeffizienten-Schätzung im dritten Teilband bei einem SNR von 5 dB.*

deutlich robuster als auf zu kleine. Besonders *Methode 3* liefert selbst bei einer nur um Eins zu klein angenommenen Ordnung ($p=4$) bereits kein brauchbares Spektrum mehr. *Methode 1* bildet in diesem Fall wenigstens noch ungefähr die wahre spektrale Struktur nach. Für $p=2$ versagen beide vollständig, und es werden nicht, wie man erwarten könnte, die zwei stärksten Pitchkomponenten abgebildet.

Für zu groß angenommene Ordnungen reagieren beide Methoden äußerst robust, weshalb die Sprachmodellordnung im Zweifel eher ein wenig zu groß gewählt werden sollte. Wird die Abweichung zu groß (im Beispiel aus Abbildung 6.6 bei $p=8$), beginnt *Methode 1*, die Frequenz der schwächsten Pitchkomponenten leicht zu verfälschen, während *Methode 3* alle Komponenten noch korrekt abbildet.

Für das in dieser Arbeit vorgeschlagene Teilbandsystem wird für das dritte Teilband eine Sprachmodellordnung von $p=6$ verwendet (siehe Tabelle 5.2). Dies entspricht bei dem hier vorgestellten Beispiel einer Überschätzung der Ordnung um Eins. Wie in Abbildung 6.6 ersichtlich, stellt dies keine Beeinträchtigung für

die Spektralschätzung dar.

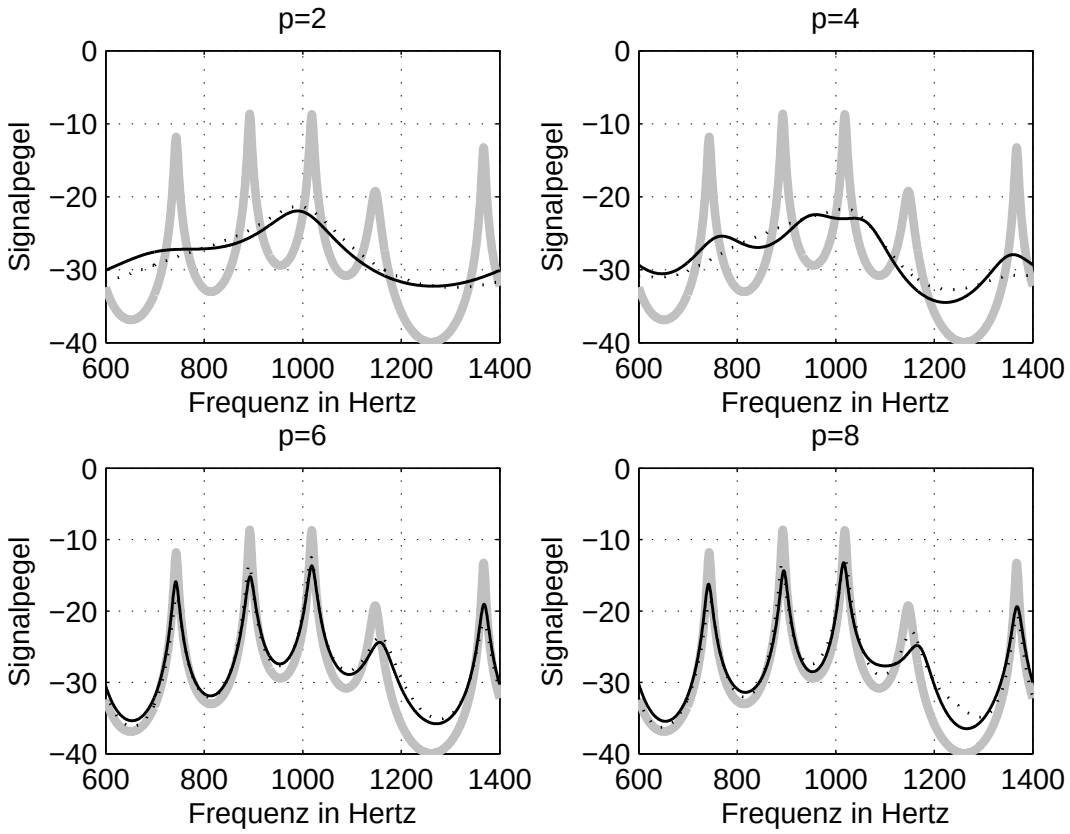

Abbildung 6.6: *Einfluss der Ordnung p des Sprachmodells auf die Güte der AR-Koeffizienten-Schätzung im dritten Teilband bei einem SNR von 15 dB.*

6.2 Leistungsfähigkeit des Gesamtsystems

Zur Bewertung der Güte von Algorithmen zur akustischen Geräuschreduktion kann die erreichte Dämpfung des Geräuschs nicht alleine als Maßstab herangezogen werden. Genauso wichtig ist, dass durch die Anwendung der Geräuschreduktion die Sprachkomponente im Ausgangssignal nicht verzerrt oder gedämpft wird. Die Bewertung der Sprachqualität wird mittels sogenannter *Mean Opinion Score (MOS)* Tests durchgeführt, bei denen eine ausreichend große Anzahl von Testpersonen die Sprachqualität auf einer Skala von Eins (sehr schlecht) bis Fünf (sehr gut) bewerten muss [52]. Die Durchführung von MOS-Tests war im Rahmen dieser Arbeit nicht möglich, so dass Aussagen zur Sprachqualität der Signale die subjektive Meinung des Autors wiederspiegeln.

Die erreichte Geräuschdämpfung wurde durch Vergleichsmessungen der Signalleistung in vorher definierten Sprachpausen bestimmt. Dazu wurden Testsignale mit längeren, von verschiedenen Sprechern gesprochenen Texten dem Geräuschreduktionsalgorithmus zugeführt, wobei zwischen den einzelnen Sequenzen eine kurze Pause, bei der nur Geräusch vorlag, eingefügt wurde. In diesen Pausen wurde jeweils vorher und nachher die Geräuschleistung bestimmt.

6.2.1 Testszenarien

Es wurden zwei Testszenarien, die im folgenden kurz beschrieben werden, untersucht und miteinander verglichen:

- MISO Kalman-Filter Struktur in verschiedenen Konfigurationen,
- Delay-and-Sum Beamformer mit anschließender einkanaliger Kalman-Filter Verarbeitung.

Unter der MISO Kalman-Filter Struktur wird das in den vorangegangenen Kapiteln beschriebene Verfahren bezeichnet. Der Delay-and-Sum Beamformer (DSBF) ist der einfachste aller Beamformer und in Abbildung 6.7 dargestellt.

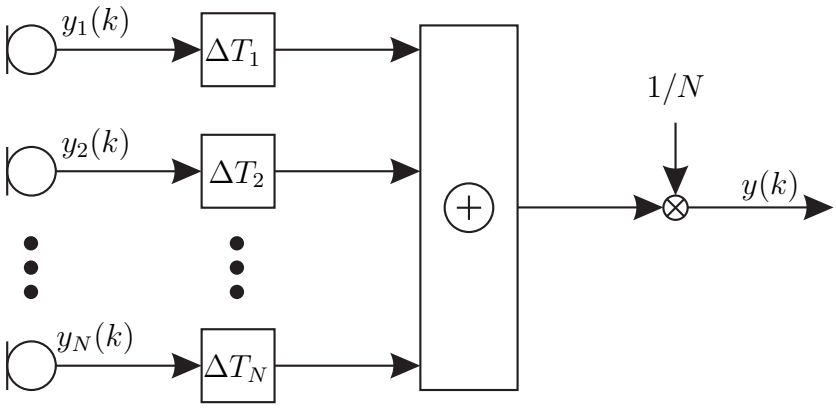

Abbildung 6.7: *Delay-and-Sum Beamformer mit Eingangssignalen $y_i(k)$, Laufzeitausgleich mit Verzögerungen ΔT_i und Ausgang $y(k)$.*

Die Mikrofonsignale $y_i(k)$ für $i = 1, \ldots, N$ werden einem Laufzeitausgleich, symbolisiert durch die Laufzeit ΔT_i (siehe Gleichung 4.57), zugeführt, addiert und auf die Mikrofonanzahl N normiert. Dadurch addieren sich Signale aus Nutzsignalrichtung konstruktiv, alle anderen Richtungen überlagern sich im Mittel destruktiv [4, 7]. Das so berechnete Ausgangssignal des Beamformers $y(k)$ wird dann anschließend dem einkanaligen Geräuschreduktionsalgorithmus zugeführt.

In Tabelle 6.1 sind die mit den unterschiedlichen Testfällen und Konfigurationen erreichten Dämpfungswerte für verschiedene Mikrofonanzahlen N (Spalten) und einem SNR von 15 dB aufgeführt. Für $N = 1$ sind alle Verfahren identisch, für $N = 2$ unterscheiden sich voll und sparsam besetzte Messmatrix $\mathbf{C}(k)$ noch nicht.

Tabelle 6.1: *Erzielbare Geräuschunterdrückung der verschiedenen Testfälle und Konfigurationen.*

Testfall / Konfiguration	$N = 1$	$N = 2$	$N = 4$	$N = 7$
MISO ($\mathbf{C}$-Matrix voll besetzt)	4,9 dB	6,1 dB	6,5 dB	6,7 dB
MISO ($\mathbf{C}$-Matrix dünn besetzt)	n/a	n/a	6,4 dB	6,7 dB
D&S BF + Kalman-Filter einkanalig	n/a	5,9 dB	6,5 dB	7,2 dB

Die Sprachqualität kann in allen Fällen als gut bezeichnet werden. Auftretende Musical Tones werden gut vom Restrauschen maskiert und sind, wenn überhaupt, nur sehr schwer zu hören. Die enthallende Wirkung ist für alle Testfälle und Konfigurationen bei $N = 2$ und $N = 4$ Mikrofonen etwa gleich. Für $N = 7$ Mikrofone weist die Beamformerstruktur leichte Vorteile auf, was auch aufgrund der etwas besseren Geräuschdämpfung zu erwarten war. Die Ein- und Ausgangssignale für den MISO Fall mit dünn besetzter $\mathbf{C}$-Matrix bei $N = 4$ und 15 dB SNR sind in den Abbildung 6.8 als Spektrogramme dargestellt.

6.3 Fazit

Die vorgeschlagenen Verfahren zur mehrkanaligen Schätzung der AR-Koeffizienten liefern wie erwartet mit steigender Kanalanzahl verbesserte Schätzwerte. Dies trifft insbesondere für sehr schlechte Signal-zu-Geräusch Verhältnisse zu. Dagegen fällt der Gewinn bei den in Kraftfahrzeugen typischerweise anzutreffenden etwas höheren SNR-Werten weniger groß aus. Dort können bereits mit wenig Kanälen ($N \leq 4$) gute Ergebnisse erzielt werden, die durch eine weitere Hinzunahme von Mikrofonen nicht mehr signifikant verbessert werden können.

Bei der Wahl der Sprachmodellordnung p ist darauf zu achten, dass diese auf keinen Fall zu klein gewählt werden darf, da bei deutlicher ($|p - p_{\mathrm{ref}}| \geq 3$) Unterschätzung der tatsächlich vorliegenden Ordnung die Gefahr besteht, dass die Schätzung vollständig versagt. Allerdings muss dabei ebenfalls beachtet werden, dass zum einen der numerische Aufwand kubisch mit der Ordnung wächst und zum anderen bei deutlich ($|p-p_{\mathrm{ref}}| \geq 3$) zu groß gewähltem p ebenfalls Schätzfehler auftreten.

Die bei der Simulation des Gesamtsystems erhaltenen Ergebnisse bestätigen die zuvor gemachten Aussagen. Auch hier kann für typische SNR-Werte und große

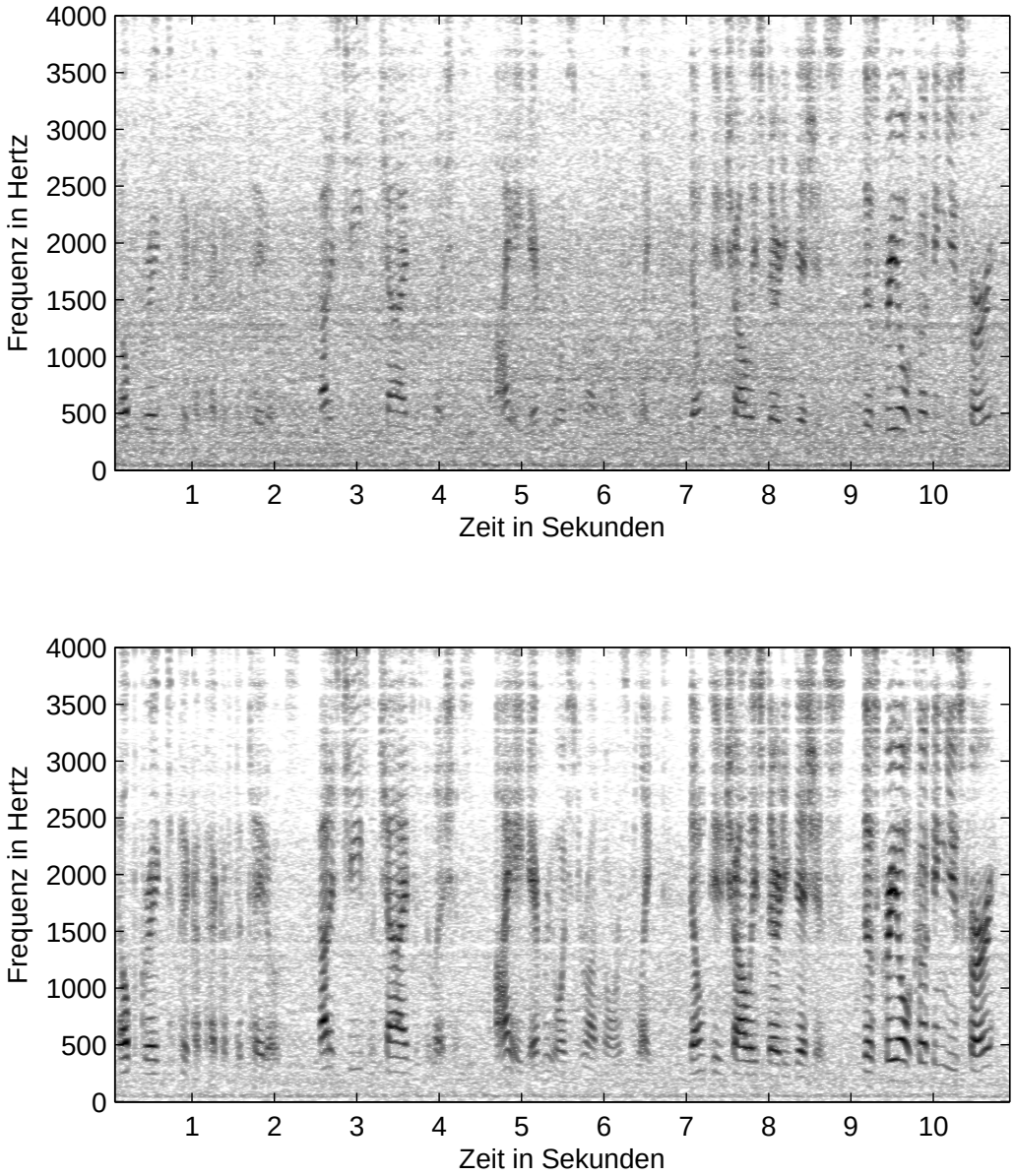

Abbildung 6.8: *Oben: Spektrogramm eines verrauschten Sprachsignals mit einem SNR von ca. 15 dB. Unten: Gleiches Signal nach Durchlaufen der Geräuschreduktion für $N = 4$ und dünn besetzter Matrix $\mathbf{C}(k)$.*

Mikrofonanzahlen $N > 4$ kaum zusätzlicher Gewinn erzielt werden. Dies liegt zum einen an der mit steigender Kanalanzahl nicht mehr signifikant besser werden-

den AR-Koeffizientenschätzung. Zum anderen kann das Kalman-Filter aufgrund der mit dem heutigen Stand der Technik nicht zuverlässig möglichen blinden Schätzung der Raumimpulsantworten den Vorteil der Mehrkanaligkeit insbesondere für viele Mikrofone nicht voll ausspielen. Deshalb beginnt die DSBF-Struktur für $N > 4$, bessere Ergebnisse zu liefern.

Aus diesem Grund erscheint mit den im Rahmen dieser Arbeit vorhandenen Möglichkeiten ein Erhöhen der Mikrofonanzahl über $N = 4$ hinaus wenig sinnvoll.

Kapitel 7

Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurden auf Kalman-Filterung basierte Verfahren zur mehrkanaligen Geräuschreduktion bei Sprachsignalen untersucht. Als Zielanwendung wurde hierfür die Freisprecheinrichtung in Kraftfahrzeugen festgelegt. Dort ist das mit den Mikrofonen aufgenommenen Sprachsignal des Sprechers durch Fahrzeuggeräusch gestört. Als Ausgangspunkt für die Entwicklung der Algorithmen wurde das in [39] vorgestellte einkanalige, auf Kalman-Filter basierende Verfahren verwendet.

Die durchgeführte Signalanalyse zeigt, dass Sprache und Fahrzeuggeräusch nicht nur den gleichen Frequenzbereich einnehmen, sondern auch eine etwa identische spektrale Grobstruktur aufweisen. Darüber hinaus konnte gezeigt werden, dass beide hinreichend genau durch AR-Prozesse modelliert werden können.

Darauf aufbauend wurde ein mehrkanaliges Signalmodell für die Herleitung des Kalman-Filters vorgestellt. In diesem werden die AR-Prozesse von Sprache und Geräusch mit den Größen des Fahrzeuginnenraums, das sind die Raumimpulsantworten und Kreuzfilter, verknüpft. Dadurch wurde eine Kombination von Beamformer- und Postfilterfunktionalität in einem Filter ermöglicht.

Auf Basis dieses im Zustandsraum formulierten Signalmodells wurde anschließend ein mehrkanaliges Kalman-Filter hergeleitet. Es wurde gezeigt, dass dessen numerische Komplexität kubisch mit der Modellordnung und der Kanalanzahl wächst. Außerdem bietet der vorgestellte Algorithmus die Möglichkeit, durch Überschätzen des Geräuschs die numerische Stabilität zu erhöhen und das Auftreten von hörbaren Musical Tones zu vermindern.

Für die notwendige Schätzung der AR-Parameter wurden drei mehrkanalige Methoden vorgeschlagen und verglichen. Kriterien waren hierbei die Differenz der geschätzten Spektren zu dem Referenzspektrum, sowie die Fähigkeit, alle im Signal vorhandenen Pitchkomponenten zu detektieren. Dabei wurde gezeigt, dass die auf Mittelung der Reflexionskoeffizienten basierten Methoden bei typischem SNR die besten Ergebnisse liefern. Dahingegen weist bei schlechtem SNR die auf gemittelten Periodogrammen basierte Methode die bessere Schätzgüte auf. Zu-

dem wurde gezeigt, dass sich alle drei Methoden robust gegenüber zu groß oder zu klein angenommenen Sprachmodellordnungen verhalten, solange die Abweichung nach unten nicht über zwei Ordnungen hinausgeht.

Als Ersatz für die blinde Schätzung der Raumimpulsantworten wurde ein Verfahren vorgeschlagen, das einen festen, auf die mittlere Sitzposition des Fahrers eingestellten, Laufzeitausgleich der Mikrofonsignale durchführt.

Durch die Implementierung in einer Teilbandstruktur wurde erreicht, dass Sprachmodellordnungen kleiner gleich sechs innerhalb der Teilbänder ausreichen, um eine vollständige Modellierung der gesamten spektralen Einhüllenden von Sprache zu ermöglichen. Dadurch wurde der numerische Aufwand gegenüber einer Implementierung im Vollband erheblich reduziert. Die Größe der Sprachmodellordnung wurde für jedes Teilband auf Grundlage der mittleren Energie und typischen Sprachgrundfrequenzen festgelegt.

Die Ergebnisse der Gesamtsystemsimulation zeigen schließlich, dass die vorgestellten Verfahren mit wachsender Mikrofonanzahl eine immer besser werdende Geräuschreduktion gegenüber der einkanaligen Methode erreichen. Für den praktischen Einsatz ist die Verwendung von vier Mikrofonen hinreichend.

Eine im Rahmen der Arbeit offen gebliebene Fragestellung ist die zuverlässige, blinde Schätzung der Raumimpulsantworten aus den verrauschten Mikrofonsignalen, um eine zur Geräuschreduktion zusätzliche Enthallung zu bewirken.

Weitere Ansatzpunkte für die zukünftige Entwicklung dieser auf Kalman-Filter basierenden Verfahren sind:

- Integration der in [39] vorgeschlagenen Methoden zur Pitch-adaptiven Verbesserung der Sprachmodellschätzung.
- Berücksichtigung der durch die Filterbankimplementierung eingeführten Nichtkausalitäten im Signalmodell.
- Automatische Verwendung der besten Schätzmethode in Abhängigkeit vom SNR des jeweiligen Teilbands.
- Untersuchung des Verhaltens der Algorithmen bei Verwendung weniger hochwertiger Mikrofone.
- Praktische Umsetzung der Verfahren in einem Kraftfahrzeug.
- Durchführung von Probanden Hörtests zur weiteren Validierung.

Notation

Wichtige Definitionen

Tabelle 7.1: *Liste wichtige Definitionen.*

Definition	*Beschreibung*
$r_{yy}(l) = \mathrm{E}\{x^*(k_1)x(k_2)\}$	Autokorrelationsfolge ($y(k)$ instationär)
$r_{xy}(k_1, k_2) = \mathrm{E}\{x^*(k)y(k+l)\}$	Kreuzkorrelationsfolge ($y(k)$ stationär)
$S_{yy}(\Omega) = \sum_{k=-\infty}^{+\infty} r_{yy}(l)e^{-j\Omega l}$	Autoleistungsdichtespektrum
$h(k) * y(k) = \sum_{l=-\infty}^{+\infty} h^*(l)y(k-l)$	Faltung zweier Folgen
$\mathbf{R}_{yy}(k)$	Autokorrelationsmatrix
$\mathbf{r}_{xy}(k)$	Kreuzkorrelationsvektor

Wichtige Abkürzungen

Tabelle 7.2: *Liste wichtiger Abkürzungen.*

Abkürzung	*Bedeutung*
ARMA	Autoregressive Moving Average
DAKF	Differenz-Autokorrelationsfunktion
(I)DFT bzw. (I)FFT	(Inverse) Diskrete bzw. Schnelle Fouriertransformation
DSBF	Delay & Sum Beamformer
LP(E)	Linearer(s) Prädiktor (Fehlerfilter)
VAD	Sprachpausendetektion

Wichtige Formelzeichen

Skalare Größen

Tabelle 7.3: *Liste wichtiger skalarer Größen (Teil 1 von 3).*

Größe	*Beschreibung*
a_i	i-ter Koeffizient des Polynoms $A(z)$
$a_{\mathrm{LP},i}$	i-ter Prädiktorkoeffizient
$a_{\mathrm{LPE},i}$	i-ter Koeffizient des Prädiktorfehlerfilters
$a_{s,i}(k)$	i-ter Prädiktorkoeffizient der Sprachkomponente
$a_{n_l,i}(k)$	i-ter Prädiktorkoeffizient der Geräuschkomponente des l-ten Mikrofons
$A(z)$	Nennerpolynom von $H_{\mathrm{ARMA}}(z)$
$A_\mu(e^{j\Omega})$	Aliasingkomponente des μ-ten Frequenzbands
b_i	i-ter Koeffizient des Polynoms $B(z)$
$B(z)$	Zählerpolynom von $H_{\mathrm{ARMA}}(z)$
$B_{\mathrm{TP}}, B_{\mathrm{HP}}, B_{\mathrm{BP}}$	Bandbreiten der Tief-, Hoch- und Bandpassfilter
B_{FB}	Effektive Bandbreite der Teilbänder
c_{Schall}	Schallgeschwindigkeit
d_i	Mittlere Distanz vom Sprecher zum i-ten Mikrofon
$e(k)$	Restfehlersignal
$e_{l_1 l_2}(k)$	Adaptionsfehler bei Einstellung von $g_{l_1 l_2}(k)$
$e^{(\mathrm{f},m)}(k)$	Vorwärts-Prädiktionsfehler der Ordnung m
$e^{(\mathrm{b},m)}(k)$	Rückwärts-Prädiktionsfehler der Ordnung m
f_0	Unter Grenzfrequenz
f_1	Obere Grenzfrequenz
f_p	Sprachgrund- bzw. Pitchfrequenz
f_s	Abtastfrequenz
g_k	Prototyptiefpassfilter der Synthesefilterbank
$g_{\mu,k}$	Tief-, Hoch- bzw. Bandpassfilter der Synthesefilterbank im μ-ten Band
$g_{l_1 l_2,i}(k)$	i-ter Koeffizient des Kreuzfilters von Mikrofon l_1 zu l_2
h_k	Prototyptiefpassfilter der Analysefilterbank
$h_{\mu,k}$	Tief-, Hoch- bzw. Bandpassfilter der Analysefilterbank im μ-ten Band
$h(k)$	Raumimpulsantwort
$h_{\mathrm{NR}}(k)$	Impulsantwort des Geräuschreduktionsfilters

Tabelle 7.4: *Liste wichtiger skalarer Größen (Teil 2 von 3).*

Größe	*Beschreibung*
$h_{l,i}(k)$	i-ter Koeffizient der Raumimpulsantwort des l-ten Mikrofons
$H_\mu(e^{j\Omega})$	Fouriertransformierte von $h_{\mu,k}$
$H_{\text{syn}}(z)$	Übertragungsfunktion des Formfilters der Sprachsynthese
$H_{\text{glot}}(z)$	Übertragungsfunktion des Glottisfilters
$H_{\text{lip}}(z)$	Übertragungsfunktion des Lippenfilters
$H_{\text{voc}}(z)$	Übertragungsfunktion des Vokaltraktfilters
$H_{\text{ARMA}}(z)$	Übertragungsfunktion des ARMA-Modells
$H_{\text{AR}}(z)$	Übertragungsfunktion des AR-Modells
$H_{\text{LP}}(z)$	Übertragungsfunktion des linearen Prädiktors
$H_{\text{LPE}}(z)$	Übertragungsfunktion des Prädiktorfehlerfilters
k	Diskrete Zeitvariable (Abtastzeitpunkt)
k'	Diskrete Zeitvariable nach Unterabtastung
K	Anzahl der Teilfolgen bei Blockzerlegung
L	Blocklänge
L_h	Länge von $h(k)$
L_K	Länge der K Teilfolgen bei Blockzerlegung
L_{AR}	Für die Parameterschätzung verwendete Blocklänge
L'_{AR}	Für die Parameterschätzung verwendete Blocklänge nach Zero-Padding
L_{PLP}	Länge der Impulsantwort des Prototyptiefpassfilters
M	Anzahl Frequenzbänder
n	Frequenzindex im DFT-Bereich
$n(k)$	Geräuschsignal
$n_i(k)$	Geräuschsignal des i-ten Mikrofons
N	Mikrofonanzahl
p	Ordnung des Sprachmodells
q	Ordnung des Geräuschmodells
Q	Blockversatz
r	Unterabtastrate
$s(k)$	Sprachsignal
$\hat{s}(k)$	geschätztes Sprachsignal
$s'(k)$	Signal am Ausgang der Sprachsynthese
$s_\eta(k)$	Sprachsignal des η-ten Signalblocks
$s_{\text{ref}}(k)$	Mit Referenzmikrofon gemessenes Sprachsignal
$S(e^{j\Omega}, \eta)$	Fouriertransformierte von $s_eta(k)$

Tabelle 7.5: *Liste wichtiger skalarer Größen (Teil 3 von 3).*

Größe	*Beschreibung*
$\hat{S}_\mu(k')$	Schätzwert der Sprachkomponente des μ-ten Frequenzbands
$SNR_i(k)$	Signal-zu-Geräuschverhältnis im i-ten Kanal
t	kontinuierliche Zeitvariable
T	Abtastintervall
ΔT_i	Laufzeit des Signals zum i-ten Mikrofon
$v(k)$	Anregungssignal des Sprachmodells
$w_i(k)$	Anregungssignal des Geräuschmodells des i-ten Mikrofons
$y(k)$	Mikrofonsignal
$y_i(k)$	Signal am Ausgang des i-ten Mikrofons
$y_{\mathrm{VAD}}(k)$	Signal des Sprachpausendetektors
$Y_\mu(k')$	Eingangsignal des μ-ten Frequenzbands
$z(k)$	zusätzliches Messrauschen
$z_l(k)$	zusätzliches Messrauschen des l-ten Mikrofons
$z_{\mathrm{P},i}$	i-te Nullstelle des Nennerpolynoms von H_{ARMA}
$z_{0,i}$	i-te Nullstelle des Zählerpolynoms von H_{ARMA}
$\breve{y}(k)$	Signal am Ausgang des Beamformers
α_{Burg}	Gewichtung des Burg-Algorithmus'
α_{DAKF}	Gewichtung der DAKF-Methode
α_{NLMS}	Schrittweite beim NLMS Algorithmus
α_n	Glättungskonstante für Geräuschglättung der DAKF
β_n	Überschätzungsfaktor für spektrale Subtraktion der DAKF
β_{SNR}	Faktor zum Einstellen des gewünschten SNRs
$\Gamma_p(k)$	Reflexionskoeffizient der p-ten Ordnung
$\hat{\Gamma}_{s_i,l}(k)$	Geschätzter Reflexionskoeffizient der l-ten Ordnung der Sprachkomponente im i-ten Mikrofon
$\hat{\Gamma}^{\mathrm{DAKF}}_{s_i,l}(k)$	Geschätzter Reflexionskoeffizient der l-ten Ordnung der Sprachkomponente im i-ten Mikrofon (DAKF)
$\hat{\Gamma}^{\mathrm{Burg}}_{s_i,l}(k)$	Geschätzter Reflexionskoeffizient der l-ten Ordnung der Sprachkomponente im i-ten Mikrofon (Burg)
η	Blockindex
μ	Frequenzbandindex
ν_{v}	Verstärkungsfaktor bei stimmhafter Anregung
ν_{uv}	Verstärkungsfaktor bei stimmloser Anregung
Ω	Normierte Kreisfrequenz
ϱ	Normierungsfaktor

Vektorielle Größen

Tabelle 7.6: *Liste wichtiger vektorieller Größen.*

Größe	*Beschreibung*
$\mathbf{a}$	Vektor der AR-Koeffizienten
$\mathbf{A}(k\|k-1)$	Transitionsmatrix
$\mathbf{A}_s(k\|k-1)$	Transitionsmatrix der Sprachkomponente
$\mathbf{A}_{n_l}(k\|k-1)$	Transitionsmatrix der Geräuschkomponente des l-ten Mikrofons
$\mathbf{B}$	In Matrix zusammengefasste Ausschneidevektoren
$\mathbf{C}(k)$	Messmatrix
$\mathbf{C}^{-}(k)$	verallgemeinerte Inverse von $\mathbf{C}(k)$
$\mathbf{C}^{+}(k)$	Pseudoinverse von $\mathbf{C}(k)$
$\mathbf{e}(k\|k)$	a-posteriori Schätzfehler
$\mathbf{e}(k\|k-1)$	a-priori Schätzfehler
$\mathbf{g}_{l_1 l_2}(k)$	Vektor der Kreuzfilterfunktion von Mikrofon l_1 zu l_2
$\mathbf{h}(k)$	Vektor der Raumimpulsantwort
$\mathbf{h}_l(k)$	Vektor der Raumimpulsantwort des l-ten Mikrofons
$\mathbf{K}(k)$	Kalman-Verstärkung
$\mathbf{n}_l(k)$	Vektor des Geräuschsignals des l-ten Mikrofons
$\mathbf{P}_{\mathbf{e}}(k\|k)$	Kovarianzmatrix des a-posteriori Schätzfehlers
$\mathbf{P}_{\mathbf{e}}(k\|k-1)$	Kovarianzmatrix des a-priori Schätzfehlers
$\mathbf{P}_{\mathbf{u}}(k)$	Kovarianzmatrix des Anregungsvektors
$\mathbf{P}_{\mathbf{x}}(k)$	Kovarianzmatrix des Zustandsvektors
$\mathbf{P}_{\mathbf{z}}(k)$	Kovarianzmatrix des Vektors des zusätzlichen Messrauschens
$\mathbf{s}(k)$	Vektor des Sprachsignals
$\mathbf{u}(k)$	Anregungsvektor
$\mathbf{x}(k)$	Zustandsvektor
$\hat{\mathbf{x}}(k\|k)$	geschätzter a-posteriori Zustandsvektor
$\hat{\mathbf{x}}(k\|k-1)$	geschätzter a-priori Zustandsvektor
$\mathbf{y}(k)$	Vektor der Mikrofonsignale bzw. Messvektor
$\hat{\mathbf{y}}(k\|k-1)$	vorhergesagter Messvektor
$\mathbf{z}(k)$	Vektor des zusätzlichen Messrauschens

Literaturverzeichnis

[1] L. Arevalo: *Beiträge zur Schätzung der Frequenzen gestörter Schwingungen kurzer Dauer und eine Anwendung auf die Analyse von Sprachsignalen*, Dissertation, Ruhr-Universität Bochum, 1993.

[2] S. Beack, S.H. Nam, M. Hahn: A New Speech Enhancement Algorithm for Car Environment Noise Cancellation with MBD and Kalman Filtering, *IEICE Trans. Fund. Electr. Comm. Computer Scien.*, vol. E88-A, no. 3, pp. 685-689, 2005.

[3] M. Bellanger, G. Bonnerot, M. Coudreuse: Digital Filtering by Polyphase Network: Application to Sampling-Rate Alteration and Filter Banks, *IEEE Trans. Acoust. Speech Signal Process.*, vol. 24, no. 2, pp. 109-114, 1976.

[4] J. Bitzer: *Mehrkanalige Geräuschunterdrückungssysteme - eine vergleichende Analyse*, Dissertation Universität Bremen, Band 9, Shaker Verlag, Herzogenrath, 2002.

[5] J. Bitzer, K.U. Simmer: Multi-Channel Speech Enhancement in a Car Environment using Wiener Filtering and Spectral Subtraction, *Proc. ICASSP '97*, vol. 2, pp. 1167-1170, Munich, Germany, 1997.

[6] R. Boite, H. Leich: A new Procedure for the Design of High-Order Minimum-Phase FIR-Digital or CCD Filters, *Signal Processing*, vol. 3, pp. 101-108, 1980.

[7] M. Brandstein, D. Ward (eds.): *Microphone Arrays*, Springer-Verlag, Heidelberg, 2001.

[8] I.N. Bronstein, K.A. Semendjajew, G. Musiol: *Taschenbuch der Mathematik*, 6. Auflage, Verlag Harri Deutsch, Frankfurt a.M., 2005.

[9] M. Brookes: *Matrix Reference Manual*, http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html, 2005.

[10] University of Colorado at Boulder: *Matrix Calculus*, Appendix D of *Introduction to Finite Element Methods*, http://www.colorado.edu.

[11] R.E. Crochiere, L.R. Rabiner: *Multirate Digital Signal Processing*, Prentice Hall Inc., Upper Saddle River, NJ, 1983.

[12] J. Dattorro: *Matrix Calculus*, appendix D of *Convex Optimization & Euclidean Distance Geometry*, http://www.stanford.edu/ dattorro/matrixcalc.pdf, Meboo Publishing, Palo Alto, CA, 2005.

[13] ETS 300 903 (GSM 03.50): *Transmission Planing Aspects of the Speech Service in the GSM Public Land Mobile Network (PLMS) System*, ETSI, France, 1999.

[14] N. Fliege: *Multiraten-Signalverarbeitung, Theorie und Anwendungen*, B. G. Teubner, Stuttgart, 1993.

[15] H. Gahlau: *Fahrzeugakustik. Entwicklung und Einsatz von Systemen zur Lärmreduzierung*, Mod. Industrie, 1998.

[16] S. Gannot, D. Burshtein, E. Weinstein: Iterative and sequential Kalman filter-based speech enhancementalgorithms, *IEEE Trans. Acoust. Speech Signal Process.*, vol. 6, no. 4, pp. 373-385, 1998.

[17] M.S. Grewal, A.P. Andrews: *Kalman Filtering – Theory and Practice using MATLAB: Theory and Practice Using MATLAB* , 2. Auflage, John Wiley & Sons, Inc., New York, NY, 2001.

[18] E. Hänsler, G.U. Schmidt: *Acoustic Echo and Noise Control – A Practical Approach*, John Wiley & Sons, Inc., New York, NY, 2004.

[19] E. Hänsler: *Statistische Signale – Grundlagen und Anwendungen*, 3. Auflage, Springer-Verlag, Heidelberg, 2001.

[20] E. Hänsler, G.U. Schmidt (eds.): *Topics in Acoustic Echo and Noise Control*, Springer-Verlag, Heidelberg, 2006.

[21] M.H. Hayes: *Statistical Digital Signal Processing and Modeling*, John Wiley & Sons, Inc., New York, NY, 1996.

[22] S. Haykin: *Adaptive Filter Theory*, 3. Auflage, Prentice-Hall, Inc., Upper Saddle River, NJ, 1996.

[23] W. Hess: *Pitch Determination of Speech Signals*, 3. Auflage, Springer-Verlag, Heidelberg, 1996.

[24] J.R. Hopgood: *Models for Blind Speech Dereverberation: A Subband All-Pole Filtered Block Stationary Autoregressive Process*, Proc EUSIPCO '05, Antaly, Turkey, 2005.

[25] R.E. Kalman: A New Approach to Linear Filtering and Prediction Problems, *Trans. ASME–Journal of Basic Engineering*, vol. 82 (Series D), pp. 35-45, 1960.

[26] K.D. Kammeyer, K. Kroschel: *Digitale Signalverarbeitung*, 4. Auflage, B. G.

Teubner, Stuttgart, 1998.

[27] A. Kaps: Acoustic Noise Reduction Using a Multiple-Input Single-Output Kalman Filter, *Proc. IWAENC '05*, pp. 133-136, Eindhoven, Netherlands, 2005.

[28] U. Kornagel: *Synthetische Spektralerweiterung von Telefonsprache*, Darmstädter Dissertation D17, Fortschr.- Ber. VDI, Reihe 10, Nr. 736, VDI Verlag, Düsseldorf, 2004.

[29] T.I. Laakso, V. Välimäki, M. Karjalainen, U.K. Laine: Splitting the Unit Delay – Tools for fractional delay filter design, *IEEE Signal Process. Magazine*, vol. 13, no. 1, pp. 30-60, 1996.

[30] N. Ma et al.: Perceptual Kalman filtering for speech enhancement in colored noise, *Proc. ICASSP '04*, vol. 1, pp. 717-720, Montreal, Canada, 2004.

[31] M. Miyoshi, Y. Kaneda: Inverse Filtering of Room Acoustics, *IEEE Trans. Acoust. Speech Signal Process.*, vol. 36, no. 2, pp. 145-152, 1988.

[32] A.V. Oppenheim, R.W. Schafer: *Zeitdiskrete Signalverarbeitung*, 3. Auflage, R. Oldenbourg Verlag, München, 1999.

[33] K.K. Paliwal, A. Basu: A speech enhancement method based on Kalman filtering, *Proc. ICASSP '87*, vol. 1, pp. 631-634, Dallas, TX, 1987.

[34] R. Penrose: A Generalized Inverse for Matrices, *Proc. Cambridge Phil. Soc.*, vol.5, pp. 406-413, 1955.

[35] M.S. Pedersen: *Matricks*, Revision 1.2, http://www2.imm.dtu.dk, 2005.

[36] K.B. Petersen, M.S. Pedersen: *The Matrix Cookbook*, Version: February 16, 2006, http://matrixcookbook.com, 2006.

[37] J.G. Proakis et al.: *Algorithms for Statistical Signal Processing*, Prentice-Hall, Inc., Upper Saddle River, NJ, 2002.

[38] H. Puder: Kalman-Filters in Subbands for Noise Reduction with Enhanced Pitch-Adaptive Speech Model Estimation, *Euro. Trans. Telecom.*, vol. 13, no. 2, pp. 139-148, 2002.

[39] H. Puder: *Geräuschreduktionsverfahren mit modellbasierten Ansätzen für Freisprecheinrichtungen in Kraftfahrzeugen*, Darmstädter Dissertation D17, Fortschr.-Ber. VDI, Reihe 10, Nr. 721, VDI Verlag, Düsseldorf, 2003.

[40] T.F. Quatieri: *Discrete-Time Speech Signal Processing*, Prentice-Hall, Inc., Upper Saddle River, NJ, 2002.

[41] C.M. Rader: An Improved Algorithm for High Speed Autocorrelation with

Application to Spectral Estimation, *IEEE Trans. Audio Electroacoust.*, vol.56, pp. 1107-1108, 1979.

[42] C.R. Rao, S.K. Mitra: *Generalized Inverse of Matrices and Its Applications*, John Wiley & Sons, Inc., New York, NY, 1971.

[43] A.H. Sayed: *Fundamentals of Adaptive Filtering*, John Wiley & Sons, Inc., New York, NY, 2003.

[44] G.U. Schmidt: *Entwurf und Realisierung eines Multiratensystems zum Freisprechen*, Darmstädter Dissertation D17, Fortschr.- Ber. VDI, Reihe 10, Nr. 674, VDI Verlag, Düsseldorf, 2001.

[45] E.G. Schukat-Talamazzini: *Automatische Spracherkennung*, Vieweg Verlag, Braunschweig, Wiesbaden, 1995.

[46] P. Stoica, R. Moses: *Introduction to Spectral Analysis*, Prentice-Hall, Inc., Upper Saddle River, NJ, 1997.

[47] Straßenverkehrsordnung (StVO): *Sonstige Pflichten des Fahrzeugführers*, Paragraph §23, Absatz 1a.

[48] J. Tilp: *Verfahren zur Verbesserung gestörter Sprachsignale unter Berücksichtigung der Grundfrequenz stimmhafter Sprachlaute*, Darmstädter Dissertation D17, Darmstadt, 2002.

[49] J. Tilp: Formant-Based Detection of Speech Distortions for a Single-Channel Spectral-Subtraction Scheme, *Proc. IWAENC '99*, pp.72-75, Pocono Manor, PA, 1999.

[50] C.C. Tseng: Design of Variable Fractional Delay Allpass Filter Using Weighted Least Squares Method, *Proc. IEEE Int. Symp. Circ. Sys.*, vol. 5, pp. 713-716, 2002.

[51] V. Välimäki, T.I. Laakso: Principles of Fractional Delay Filters, *Proc. ICASSP '00*, vol.6, pp. 3870-3873, Istanbul, Turkey, 2000.

[52] P. Vary, W. Hess, U. Heute: *Digitale Sprachsignalverarbeitung*, B. G. Teubner, Stuttgart, 1998.

[53] P. Vary, R. Martin: *Digital Speech Transmission*, John Wiley & Sons, Inc., New York, NY, 2006.

[54] P. Vary, G. Wackersreuther: A Unified Approach to Digital Polyphase Filter Banks, *AEÜ*, Band 37, Heft 1/2, pp. 387-400, 1983.

[55] G. Wackersreuther: On the Design of Filters for Ideal QMF and Polyphase Filter Banks, *AEÜ*, Band 39, Heft 2, pp. 123-130, 1985.

[56] C. Wagner: *Untersuchung mehrkanaliger Verfahren zur Sprachpausendetektion und Hintergrundgeräuschschätzung*, Studienarbeit, TU Darmstadt, Fachgebiet Theorie der Signale, 2004.

[57] T. Wolf: *Implementierung und Untersuchung einer Generalized Sidelobe Canceller Struktur für die Verarbeitung verrauschter Sprachsignale mittels Mikrofonarray*, Studienarbeit, TU Darmstadt, Fachgebiet Theorie der Signale, 2004.

[58] W.R. Wu, P.C. Chen: Subband Kalman filtering for speech enhancement, *IEEE Trans. Circ. Syst. II*, vol. 45, no. 8, pp. 1072-1083, 1998.

[59] C.H. You, S.N. Koh, S. Rahardja: Subband Kalman filtering incorporating masking properties for noisy speech signal, *Speech Communication*, vol. 49, no. 7-8, pp. 558-573, 2007.

[60] E. Zwicker: *Psychoakustik*, Springer-Verlag, Heidelberg, 1982.